



Munich Personal RePEc Archive

A test for model specification of diffusion processes

Song Xi Chen and Jiti Gao and Chenghong Tang

Iowa State University, The University of Adelaide, Iowa State University

November 2005

Online at <http://mpra.ub.uni-muenchen.de/11976/>

MPRA Paper No. 11976, posted 9. December 2008 00:25 UTC

A Test for Model Specification of Diffusion Processes

Song Xi Chen and Jiti Gao

Iowa State University and The University of Western Australia

Abstract: This paper evaluates the use of the nonparametric kernel method for testing specification of diffusion models as originally considered in Aït-Sahalia (1996). A serious doubt on the ability of the kernel method for diffusion model testing has been cast in Pritsker (1998), who observes severe size distortion of the test proposed by Aït-Sahalia and finds that 2755 years of data are required in order for the kernel density estimator to attain a level of accuracy achieved with 22 years of independent data. We introduce in this paper a set of measures to formulate a new test based on the kernel method and show that the severe size distortion observed by Pritsker (1998) can be overcome. The measures include targeting at the transitional density of the process, using the empirical likelihood to formulate the test statistics, properly smoothing of model-implied transitional densities and employing a parametric bootstrap procedure in approximating the distribution of test statistics. Our simulation for both the Vasicek and Cox-Ingersoll-Ross diffusion models indicates that the proposed test has reasonable size and power under various degrees of data dependence for as little as 10 years of data. We then apply the proposed test to a monthly Federal Fund rate data and find that there is some empirical support for several of the one-factor diffusion models proposed in the literature.

KEYWORDS: Bootstrap, Continuous-time interest rate model, Diffusion process, Empirical likelihood, Goodness-of-fit test, Kernel method.

1. Introduction. Let X_1, \dots, X_{n+1} be $n + 1$ equally spaced (with spacing Δ in time) observations of a diffusion process

$$dX_t = \mu(X_t)dt + \sigma(X_t)dB_t \quad (1.1)$$

where $\mu(\cdot)$ and $\sigma^2(\cdot) > 0$ are, respectively, the drift and diffusion functions, and B_t is the standard Brownian motion. Suppose a parametric specification of model (1.1) is

$$dX_t = \mu(X_t; \theta)dt + \sigma(X_t; \theta)dB_t, \quad (1.2)$$

where θ is a parameter within a parameter space $\Theta \subset R^d$ for a positive integer d . The focus of this paper is testing the validity of the parametric specification (1.2) based on the discretely observed data $\{X_t\}_{t=1}^{n+1}$.

In a pioneer work that represents a major break-through in financial econometrics, Aït-Sahalia (1996) proposes using the nonparametric kernel method to test the parametric specification (1.2). The test statistic is a L_2 -distance between the kernel stationary density estimator and the stationary density implied by model (1.2). The test is based on the asymptotic normal distribution of the test statistic.

In a comprehensive investigation, Pritsker (1998) evaluates Aït-Sahalia's test and finds a large discrepancy between the simulated and nominal sizes of the test under a set of Vasicek (1977) diffusion models with various degrees of dependence. He observes that both the form of the density estimator and its asymptotic distribution are the same for both independent and dependent observations, which is due to the so-called 'pre-whitening' by a smoothing bandwidth in nonparametric curve estimation. Hence, the asymptotic nature of the Aït-Sahalia's test prevents capturing dependence existed in the process. Pritsker shows that to attain the same level of accuracy for the kernel estimator with 22 years of independent data would require 2755 years of dependent data under the Vasicek model. In a separate study, Chapman and Pearson (2000) reveal that the finding of non-linear drift in Stanton (1997) is largely due to

the boundary and truncation bias associated with the kernel regression estimator. These findings have illustrated the challenge when applying the kernel estimator for inference of diffusion processes, and have inevitably projected rather negatively on the ability of the kernel method to model data dependence induced by a diffusion process in general and testing of the diffusion process in particular.

We think other aspects contribute to the performance of Aït-Sahalia's test in addition to its asymptotic nature observed by Pritsker. One is that the test targeted at the stationary density. It can take long time for a process to settle on the stationary distribution which is specially the case for the Vasicek models with weak mean-reversion considered in Pritsker's simulation. As pointed out in Aït-Sahalia (1996), a test on the stationary density is not conclusive as two different processes may share the same stationary density. As diffusion processes are Markovian, testing should be aimed at the transitional density. The second aspect is that the kernel estimator introduces a bias which needs to be considered. Aït-Sahalia proposed undersmoothing to prevent the bias from getting into the limiting distribution. However, it may not be easy to check on the effect of a particular bandwidth used on the bias. Another aspect is that parameter estimates under the hypothesized process are needed in the formulation of a test statistic. However, the maximum likelihood estimators for the drift parameters are subject to large bias when the process is weak mean-reverting. This is a well-known problem in diffusion process estimation (Yu and Phillips, 2001) and further reduces the performance of Aït-Sahalia's test.

We implement a set of measures in developing a new test based on the kernel method. First of all, the proposed test is targeted at the transitional density to have a conclusive test and to fully capture the dynamics. The second measure is to properly smooth the model implied transitional density so as to cancel out the bias induced by the kernel estimation, which avoids undersmoothing and simplifies

the theoretical analysis. In order to put the difference between the kernel estimator and the smoothed model-implied transitional density in the context of its variation, the test statistic is formulated via the empirical likelihood of Owen (1988, 1990). Furthermore, to make the test robust against bandwidth choice, we formulate the test statistic based a set of bandwidths. Finally, a parametric bootstrap procedure is used to profile the distribution of the test statistic and to obtain the critical value of the test.

A continuous-time diffusion process and a time series model share some important features although one is a continuous time model and the other is a discrete model. Time series and diffusion processes can be both Markovian and weakly dependent satisfying certain mixing conditions. Kernel-based tests have been shown to be able to effectively test for discrete time series models as demonstrated in Robinson (1989), Hjellvik and Tjøstheim (1995), Fan and Li (1996), Hjellvik and Yao and Tjøstheim (1998), Kreiss, Neumann and Yao (1998), Li (1999), Aït-Sahalia, Bickel and Stoker (2001), Gozalo and Linton (2001). See the books by Hart (1997) and Fan and Yao (2003) for extended lists of references. For estimation of diffusion models, in addition to Aït-Sahalia (1996) and Stanton (1997), Jiang and Knight (1997) proposed a semiparametric kernel estimator for the drift. Fan, Jiang, Zhang and Zhou (2003) examined effects of high order stochastic expansions. Bandi and Phillips (2003) consider a two-stage kernel estimation procedure for the drift and diffusion functions without the strictly stationary assumption. For testing of diffusion processes, Hong and Li (2005) developed a test via a conditional probability integral transformation. Although the kernel estimator is employed, the transformation leads to asymptotically independent uniform random variables under the correct model. Hence, the issue on the ability of the kernel method to model dependence induced by diffusion models is avoided. Fan and Zhang (2003) apply the generalized likelihood ratio test

for the drift and diffusion functions. See review articles by Cai and Hong (2003) and Fan (2005) for more detailed accounts.

The paper is structured as follows. Section 2 outlines the hypotheses and the kernel estimation of the transitional densities. The proposed EL test is given in Section 3. Section 4 reports main results of the test. Section 5 considers computational issues. Results from simulation studies are reported in Section 6. A Federal fund rate data is analyzed in Section 7. All the proofs are given in the appendix.

2. The hypotheses and kernel estimators. Let $\pi(x)$ be the stationary density and $p(y|x; \Delta)$ be the transitional density of $X_{t+1} = y$ given $X_t = x$ under model (1.1) respectively; and $\pi_\theta(x)$ and $p_\theta(y|x, \Delta)$ be their parametric counterparts under model (1.2). To simplify notation, we suppress Δ in the notation of the transitional densities. Let \mathcal{X} be the state space of the process.

Although $\pi_\theta(x)$ has a close form expression via Kolmogorov forward equation

$$\pi_\theta(x) = \frac{\xi(\theta)}{\sigma^2(x, \theta)} \exp \left\{ \int_{x_0}^x \frac{2\mu(t, \theta)}{\sigma^2(t, \theta)} dt \right\},$$

where $\xi(\theta)$ is a normalizing constant, $p_\theta(y|x)$ implicitly defined by the Kolmogorov backward-equation may not admit a close form expression. However, this problem is overcome by Edgeworth type approximations developed by Aït-Sahalia (1999, 2002) for general diffusion processes. As the transitional density fully describes the dynamics of a diffusion process, the hypotheses we would like to test are

$$\begin{aligned} H_0 : p(y|x) &= p_{\theta_0}(y|x) \text{ for some } \theta_0 \in \Theta \text{ and all } (x, y) \in S \subset \mathcal{X}^2 \quad \text{versus} \\ H_1 : p(y|x) &\neq p_\theta(y|x) \text{ for all } \theta \in \Theta \text{ and some } (x, y) \in S \subset \mathcal{X}^2, \end{aligned} \quad (2.1)$$

where S is a compact set within \mathcal{X}^2 and can be chosen upon given observations as demonstrated in simulation and case studies in Sections 6 and 7. As we choose S within the support of the density, the boundary bias associated with the kernel

estimators (Müller, 1991; Fan and Gijbels, 1996; and Müller and Stadtmüller, 1999) is avoided.

Let $K(\cdot)$ be a kernel function which is a symmetric probability density function, h is a smoothing bandwidth such that $h \rightarrow 0$ and $nh^2 \rightarrow \infty$ as $n \rightarrow \infty$, and $K_h(\cdot) = h^{-1}K(\cdot/h)$. The kernel estimator of $p(y|x)$ is

$$\hat{p}(y|x) = n^{-1} \sum_{t=1}^n K_h(x - X_t) K_h(y - X_{t+1}) / \hat{\pi}(x). \quad (2.2)$$

where $\hat{\pi}(x) = (n+1)^{-1} \sum_{t=1}^{n+1} K_h(x - X_t)$ is the kernel estimator of the stationary density used in Aït-Sahalia(1996). The local polynomial estimator introduced by Fan, Yao and Tong (1996) can be also employed without altering the main results of the paper. It is known (Hydman and Yao, 2002) that

$$E\{\hat{p}(y|x) - p(y|x)\} = \frac{1}{2} \sigma_k^2 h^2 \left(\frac{\partial^2 p(y|x)}{\partial x^2} + \frac{\partial^2 p(y|x)}{\partial y^2} + 2 \frac{\pi'(x)}{\pi(x)} \frac{\partial p(y|x)}{\partial x} \right) + o(h^2) \quad \text{and} \quad (2.3)$$

$$\text{Var}\{\hat{p}(y|x)\} = \frac{R^2(K)p(y|x)}{nh^2\pi(x)}(1 + o(1)), \quad (2.4)$$

where $\sigma_K^2 = \int u^2 K(u) du$ and $R(K) = \int K^2(u) du$.

Let $\tilde{\theta}$ be a \sqrt{n} -consistent estimator of θ under model (1.2) for instance the maximum likelihood estimator under H_0 , and

$$w_t(x) = K_h(x - X_t) \frac{s_{2h}(x) - s_{1h}(x)(x - X_t)}{s_{2h}(x)s_{0h}(x) - s_{1h}^2(x)} \quad (2.5)$$

be the local linear weight with $s_{rh}(x) = \sum_{s=1}^n K_h(x - X_s)(x - X_s)^l$ for $r = 0, 1$ and 2. In order to cancel the bias in $\hat{p}(y|x)$, we smooth $p_{\tilde{\theta}}(y|x)$ as

$$\tilde{p}_{\tilde{\theta}}(y|x) = \frac{\sum_{t=1}^{n+1} K_h(x - X_t) \sum_{s=1}^{n+1} w_s(y) p_{\tilde{\theta}}(X_s|X_t)}{\sum_{t=1}^{n+1} K_h(x - X_t)}. \quad (2.6)$$

Here we apply the kernel smoothing twice: first for each X_t using the local linear weight to smooth $p_{\tilde{\theta}}(X_s|X_t)$ and then employing the standard kernel weight to smooth the resulting function with respect to X_t . This is motivated by Härdle and Mammen

(1993). It can be shown following the standard derivations as those demonstrated in Fan and Gijbels (1996) that, under H_0

$$E\{\hat{p}(y|x) - \tilde{p}_{\hat{\theta}}(y|x)\} = o(h^2) \quad \text{and} \quad (2.7)$$

$$\text{Var}\{\hat{p}(y|x) - \tilde{p}_{\hat{\theta}}(y|x)\} = \text{Var}\{\hat{p}(y|x)\}\{1 + o(1)\}. \quad (2.8)$$

Hence the bias of $\hat{p}(y|x)$ cancels that of $\tilde{p}_{\hat{\theta}}(y|x)$ in the leading order whileas smoothing the parametric density does not effect the asymptotic variance.

3. Formulation of a test statistic. The test statistic is formulated by the empirical likelihood (EL) (Owen, 1988, 1990). Despite its being intrinsically non-parametric, EL posesses two key properties of a parametric likelihood: the Wilks' theorem and Bartlett correction. Qin and Lawless (1994) established EL for parameters defined by generalized estimating equations which is the broadest framework for EL formulation, which was extended by Kitamura (1997) to dependent observations. Chen and Cui (2004) show that the EL admits Bartlett correction under this general framework. An extension of Qin and Lawless's framework is given in Hjort, McKeague and Van Keilegom (2005) to include nuisance parameters/functions. Fan and Zhang (2004) propose a sieve EL test for varying-coefficient regression model that extends the test of Fan, Zhang and Zhang (2001). Tripathi and Kitamura (2003) propose an EL test for conditional moment restrictions. See also Li and Van Keilegom (2002) and Li (2003) for EL goodness-of-fit tests for survival data.

Let us now formulate the EL for the transitional density at a fixed (x, y) . For $t = 1, \dots, n$, let $q_t(x, y)$ be non-negative weights allocated to (X_t, X_{t+1}) . The EL evaluated at $\tilde{p}_{\hat{\theta}}(y|x)$ is

$$L\{\tilde{p}_{\hat{\theta}}(y|x)\} = \max \prod_{t=1}^n q_t(x, y) \quad (3.1)$$

subject to $\sum_{t=1}^n q_t(x, y) = 1$ and

$$\sum_{t=1}^n q_t(x, y) K_h(x - X_t) K_h(y - X_{t+1}) = \tilde{p}_{\hat{\theta}}(y|x) \hat{\pi}(x). \quad (3.2)$$

By introducing a Lagrange multiplier $\lambda(x, y)$, the optimal weights as solutions to (3.1) and (3.2) are

$$q_t(x, y) = n^{-1} \{1 + \lambda(x, y) T_t(x, y)\}^{-1}, \quad (3.3)$$

where $T_t(x, y) = K_h(x - X_t)K_h(y - X_{t+1}) - \tilde{p}_{\hat{\theta}}(x, y)$ and $\lambda(x, y)$ is the root of

$$\sum_{t=1}^n \frac{T_t(x, y)}{1 + \lambda(x, y) T_t(x, y)} = 0. \quad (3.4)$$

The overall maximum EL is achieved at $q_t(x, y) = n^{-1}$ which maximize (3.1) without the constraint (3.2). Hence, the log-EL ratio is

$$\ell\{\tilde{p}_{\hat{\theta}}(y|x)\} = -2 \log([L\{\tilde{p}_{\hat{\theta}}(y|x)\}n^n]) = 2 \sum \log\{1 + \lambda(x, y) T_t(x, y)\}. \quad (3.5)$$

It may be shown by similar derivations to those given in Chen, Härdle and Li (2003) that

$$\sup_{(x, y) \in S} |\lambda(x, y)| = o_p\{(nh^2)^{-1/2} \log(n)\}. \quad (3.6)$$

Let $\bar{U}_1(x, y) = (nh^2)^{-1} \sum T_t(x, y)$ and $\bar{U}_2(x, y) = (nh^2)^{-1} \sum T_t^2(x, y)$. From (3.4) and (3.6), $\lambda(x, y) = \bar{U}_1(x, y) \bar{U}_2^{-1}(x, y) + O_p\{(nh^2)^{-1} \log^2(n)\}$ uniformly with respect to $(x, y) \in S$. This then leads to

$$\begin{aligned} \ell\{\tilde{p}_{\hat{\theta}}(y|x)\} &= nh^2 \bar{U}_1^2(x, y) \bar{U}_2^{-1}(x, y) + O_p\{n^{-1/2} h^{-1/2} \log^3(n)\} \\ &= nh^2 \frac{\{\hat{p}(y|x) - \tilde{p}_{\hat{\theta}}(y|x)\}^2}{V(y|x)} + O_p\{h^2 + n^{-1/2} h^{-1/2} \log^3(n)\} \end{aligned} \quad (3.7)$$

uniformly for $(x, y) \in S$, where $V(y|x) = R^2(K)p(y|x)\pi^{-1}(x)$. Hence, the EL ratio is a studentized local goodness-of-fit measure between $\hat{p}(y|x)$ and $\tilde{p}_{\hat{\theta}}(y|x)$ as $\text{Var}\{\hat{p}(y|x)\} = \frac{V(y|x)}{nh^2}$.

Integrating the EL ratio with a weight function $\omega(\cdot, \cdot)$ supported on S , the global goodness-of-fit measure based on a single bandwidth is

$$N(h) = \int \int \ell\{\tilde{p}_{\hat{\theta}}(y|x)\} \omega(x, y) dx dy. \quad (3.8)$$

To make the test less dependent on a single h , we extend $N(h)$ over a bandwidth set $\mathcal{H} = \{h_k\}_{k=1}^J$ where $h_{k+1}/h_k = a$ for some $a \in (0, 1)$, whose choice can be guided by the cross-validation method of Fan and Yim (2005). This formulation is motivated by Horowitz and Spokoiny (2001) who considered achieving the optimal convergence rate of the local alternative hypothesis in testing regression models. A similar approach was applied in Fan (1996) using wavelets.

The final test statistic that bases on the bandwidth set \mathcal{H} is

$$L_n = \max_{1 \leq k \leq J} \frac{N(h) - 1}{\sqrt{2h}}.$$

The standardization reflects that $\text{Var}[N(h)] = O(2h^2)$ as elaborated in the appendix.

4. Main Results. This section establish both asymptotic distribution and asymptotic consistency for L_n . In order to state the main results, we need to introduce the following conditions.

Assumption 1. (i) Assume that the process $\{X_t\}$ is strictly stationary and α -mixing with the mixing coefficient $\alpha(t) = C_\alpha \alpha^t$ defined by $\alpha(t) = \sup\{|P(A \cap B) - P(A)P(B)| : A \in \Omega_1^s, B \in \Omega_{s+t}^\infty\}$ for all $s, t \geq 1$, where $0 < C_\alpha < \infty$ and $0 < \alpha < 1$ are constants, and Ω_i^j denotes the σ -field generated by $\{X_t : i \leq t \leq j\}$.

(ii) $K(\cdot)$ is a bounded symmetric probability density supported on $[-1, 1]$ and has bounded second derivative; and let $\sigma_K^2 =: \int_{-\infty}^{\infty} x^2 K(x) dx$ and $R(K) =: \int_{-\infty}^{\infty} K^2(x) dx$.

(iii) The bandwidth set $\mathcal{H} = \{h_k = h_{\max} a^k, k = 0, 1, 2, \dots, J\}$, where $0 < a < 1$, $h_1 = c_{\min} n^{-\gamma_1}$ and $h_{\max} = h_J = c_{\max} n^{-\gamma_2}$, in which $\frac{1}{7} < \gamma_2 \leq \gamma_1 < \frac{1}{2}$, c_{\min} and c_{\max} are constants satisfying $0 < c_{\min}, c_{\max} < \infty$, and J is an integer not depending on n .

(iv) $\omega(x, y)$ is a bounded probability density supported on S .

Assumption 2. (i) Each of the diffusion processes given in (1.1) and (1.2) admits a unique weak solution such that the boundaries of \mathcal{X} is not attainable in finite

time. In addition, each of the processes possesses a transitional density with $p(y|x) = p(y|x, \Delta)$ for model (1.1) and $p_\theta(y|x) = p_\theta(y|x, \Delta)$ for model (1.2), respectively.

(ii) Let $p_{s_1, s_2, \dots, s_l}(\cdot)$ be the joint probability density of $(X_{1+s_1}, \dots, X_{1+s_l})$ ($1 \leq l \leq 6$). Assume that each $p_{s_1, \tau_2, \dots, s_l}(x)$ is three times differentiable in $x \in \mathcal{X}^l$ for $1 \leq l \leq 6$.

(iii) The parameter space Θ is an open subset of R^d and $p_\theta(y|x)$ is three times differentiable in $\theta \in \Theta$. For every $\theta \in \Theta$, $\mu(x; \theta)$ and $\sigma^2(x; \theta)$, and $\mu(x)$ and $\sigma^2(x)$ are all three times continuously differentiable in $x \in \mathcal{X}$, and both $\sigma(x)$ and $\sigma(x; \theta)$ are positive for $x \in S$ and $\theta \in \Theta$.

Assumption 3. (i) There is a positive and integrable function $G(x, y)$ satisfying $E[\max_{1 \leq t \leq n} G(X_t, X_{t+1})] < \infty$ uniformly in $n \geq 1$ such that $\sup_{\theta \in \Theta} |p_\theta(y|x)|^2 \leq G(x, y)$ and $\sup_{\theta \in \Theta} \|\nabla_\theta^j p_\theta(y|x)\|^2 \leq G(x, y)$ for $j = 1, 2, 3$, where $\nabla_\theta p_\theta(\cdot|\cdot) = \frac{\partial p_\theta(\cdot|\cdot)}{\partial \theta}$, $\nabla_\theta^2 p_\theta(\cdot|\cdot) = \frac{\partial^2 p_\theta(\cdot|\cdot)}{(\partial \theta)^2}$ and $\nabla_\theta^3 p_\theta(\cdot|\cdot) = \frac{\partial^3 p_\theta(\cdot|\cdot)}{(\partial \theta)^3}$.

(ii) $p(y|x) > c_1 > 0$ for all $(x, y) \in S$ and that the stationary density $\pi(x) > c_2 > 0$ for all $x \in S_x$ which is the projection of S on \mathcal{X} for some $c_i > 0$ $i = 1, 2$.

(iii) there is a finite $C > 0$ such that for every $\varepsilon > 0$

$$E \left(\inf_{\theta, \theta' \in \Theta: \|\theta - \theta'\| \geq \varepsilon} [p_\theta(X_{t+1}|X_t) - p_{\theta'}(X_{t+1}|X_t)]^2 \right) \geq C\varepsilon^2.$$

Assumption 4. (i) Let H_0 be true. There exists a $\theta_0 \in \Theta$ such that $E[|\tilde{\theta} - \theta_0|^2] \leq C_{1L}n^{-1}$ for all sufficiently large n and a suitable constant $C_{1L} > 0$.

(ii) Let H_1 be true. Then there is a $\theta_1 \in \Theta$ such that $E[|\tilde{\theta} - \theta^*|^2] \leq C_{2L}n^{-1}$ for all sufficiently large n and a suitable constant $C_{1L} > 0$.

Assumption 1(i) imposes the strict stationarity and the α -mixing condition on $\{X_t\}$. Under certain conditions, such as Assumption A2 of Aït-Sahalia (1996) and Conditions (A4) and (A5) of Genon-Catalot, Jeantheau and Larédo (2000), Assumption 1(i) holds. Assumption 1(ii)(iii) is a quite standard condition imposed in kernel estimation. Assumption 2 is needed to ensure the existence and uniqueness of a solution and the transitional density function of the diffusion process. Such an assumption

may be implied under Assumptions 1–3 of Aït-Sahalia (2002), which also cover non-stationary cases. For the stationary case, Assumptions A0 and A1 of Aït-Sahalia (1996) ensure the existence and uniqueness of a stationary solution of the diffusion process. Assumption 3 imposes additional conditions to ensure the smoothness of the transitional density and the identifiability of the parametric transitional density. Assumption 4(i) requires the usual rate of convergence for $\tilde{\theta}$ to θ_0 . Such a rate is achievable when $\tilde{\theta}$ is the maximum likelihood estimator. The θ^* in Assumption 4(ii) can be regarded as a projection of the parameter estimator $\tilde{\theta}$ under \mathcal{H}_1 onto the null parameter space.

Let $K^{(2)}(z, c) = \int K(u)K(z + cu)du$ be a generalization to the convolution of K , $\nu(t) = \int \{K^{(2)}(tu, t)\}^2 du \int \{K^{(2)}(v, t)\}^2 dv$ and $\Sigma_J = \frac{2}{R^4(K)} \int \int \omega^2(x, y) dx dy (\nu(a^{i-j}))_{J \times J}$ be a $J \times J$ matrix where a is the fixed factor used in the construction of \mathcal{H} . Furthermore, Let 1_J be a J -dimensional vector of ones and $\beta = \frac{1}{C(K)R(K)} \int \int \frac{p(x, y)}{\pi(y)} \omega(x, y) dx dy$.

We now state the following asymptotic normality; its proof is given in Appendix A.

Theorem 1. *Under Assumptions 1–4 and H_0 , $L_n \xrightarrow{d} \max_{1 \leq k \leq J} Z_k$ as $n \rightarrow \infty$ where $Z = (Z_1, \dots, Z_J)^T \sim N(\beta 1_J, \Sigma_J)$.*

Theorem 1 brings a little surprise in that the mean of Z is non-zero. This is because, although the variance of $\tilde{p}_\theta(x, y)$ is at a smaller order than that of $\hat{p}(x, y)$, it contributes to the second order mean of $N(h)$ which emerges after the standardization. However, this will not affect the test as shown in Theorem 2.

We are reluctant to formulate a test based on Theorem 1 as such a test would be slow converging too. Instead we propose the following parametric bootstrap procedure to approximate l_α , the $1 - \alpha$ quantile of L_n where $\alpha \in (0, 1)$ is the nominal significance level:

Step 1: Generate an initial value X_0^* from the estimated stationary density $\pi_{\tilde{\theta}}(\cdot)$. Then simulate a sample path $\{X_t^*\}_{t=1}^{n+1}$ at the same frequency Δ according to $dX_t = \mu(X_t; \tilde{\theta})dt + \sigma(X_t; \tilde{\theta})dB_t$.

Step 2: Let $\tilde{\theta}^*$ be the estimate of θ based on $\{X_t^*\}_{t=1}^{n+1}$. Compute the test statistic L_n based on the resampled path and denote it by L_n^* .

Step 3: For a large positive integer B , repeat Steps 1 and 2 B times and obtain after ranking $L_n^{1*} \leq L_n^{2*} \leq \dots \leq L_n^{B*}$.

Let l_α^* be the $1 - \alpha$ quantile of L_n^* satisfying $P(L_n^* \geq l_\alpha^* | \{X_t\}_{t=1}^{n+1}) = \alpha$. A Monte Carlo approximation of l_α^* is $L_n^{[B(1-\alpha)]+1*}$. The proposed test rejects H_0 if $L_n \geq l_\alpha^*$.

The next theorem shows that the proposed EL test based on the bootstrap calibration has correct size asymptotically and is consistent. The proof is given in Appendix A.

Theorem 2. *Under Assumptions 1–4, then $\lim_{n \rightarrow \infty} P(L_n \geq l_\alpha^*) = \alpha$ under H_0 ; and $\lim_{n \rightarrow \infty} P(L_n \geq l_\alpha^*) = 1$ under H_1 .*

5. Computation. The computation of the test statistic L_n involves first computing the EL ratio $\ell\{\tilde{p}_{\hat{\theta}}(y|x)\}$ over a grid of (x, y) -points within the set $S \subset \mathcal{X}^2$. Then, $N(h)$ in (3.8) is computed by its Riemann sum over a set of grid points in S upon given the weight function $\omega(\cdot, \cdot)$. The EL test statistic L_n is obtained by taking the maximum of $N(h)$ over $h \in \mathcal{H}_n$. The critical value l_α is approximated via the bootstrap procedure.

Despite being computationally intensive in each of these steps, implementing the proposed test for a single data set is not a problem with a standard computing capacity these days. However, when carrying out simulations, we would like to speed up the computation as a large number of replications are required.

In the simulation, we use the least squares empirical likelihood (LSEL) to replace the full EL when formulating $N(h)$. The LSEL is easier to compute as there are

closed-form solutions for the weights $q_t(x, y)$ and hence avoids expensive nonlinear optimization. The log LSEL ratio evaluated at $\tilde{p}_{\hat{\theta}}(y|x)$ is

$$\text{lsl}\{\tilde{p}_{\hat{\theta}}(y|x)\} = \min \sum_{t=1}^n \{nq_t(x, y) - 1\}^2$$

subject to $\sum_{t=1}^n q_t(x, y) = 1$ and $\sum_{t=1}^n q_t(x, y)T_t(x, y) = 0$. According to Brown and Chen (1998), the LSEL weights are given by

$$q_t(x, y) = n^{-1} + \{n^{-1}T(x, y) - T_t(x, y)\}^T S^{-1}(x)T(x, y)$$

where $T(x, y) = \sum_{t=1}^n T_t(x, y)$ and $S(x, y) = n^{-1} \sum_{t=1}^n T_t^2(x, y)$. And

$$\text{lsl}\{\tilde{m}_{\hat{\theta}}(x)\} = S^{-1}(x, y)T^2(x, y)$$

is readily computable. The LSEL counterpart to $N(h)$ is $N^{ls}(h) = \int \text{lsl}\{\tilde{m}_{\hat{\theta}}(x)\}\pi(x, y)dx$.

The final test statistic $L_n = \max_{h \in \mathcal{H}} \frac{N^{ls}(h)-1}{C(K)h}$. It can be shown from Brown and Chen (1998) that $N^{ls}(h)$ and $N(h)$ are equivalent to the first order. Therefore, those first-order results in Theorems 1 and 2 continue to hold for the LSEL.

6. Simulation studies. We report results of simulation studies which designed to evaluate the empirical performance of the proposed EL test. To gain information on its relative performance, Hong and Li's test is performed over the same simulation.

Throughout the paper, the biweight kernel $K(u) = \frac{15}{16}(1 - u^2)^2 I(|u| \leq 1)$ was used in all the kernel estimation. In simulation, we set $\Delta = \frac{1}{12}$ implying monthly observations which coincides with that of the Federal fund rate data to be analyzed. We chose $n = 125, 250$ or 500 respectively corresponding roughly 10 to 40 years of data. The number of simulations was 500 and the number of bootstrap resampled paths was $B = 250$.

6.1. Size Evaluation.

Two simulation studies were carried out to evaluate the size of the proposed test. The first one includes three Vasicek models from Pritsker's study, whereas in the

second study three Cox, Ingersoll and Ross (CIR) (1985) models are considered. We want to see if the severe size distortion observed by Pritsker (1998) is present for our proposal.

6.1.1 Vasicek Models

We first consider, like Pritsker (1998), testing the Vasicek model

$$dX_t = \kappa(\alpha - X_t)dt + \sigma dB_t.$$

The vector of parameters $\theta = (\alpha, \kappa, \sigma^2)$ takes three sets of values which correspond to Model -2, Model 0 and Model 2 of Pritsker (1998). The baseline Model 0 assigns $\kappa_0 = 0.85837$, $\alpha_0 = 0.089102$ and $\sigma_0^2 = 0.0021854$ which matches estimates by Aït-Sahalia (1996) for an interest rate data. Model -2 is obtained by quadrupling k_0 and σ_0^2 and Model 2 by halving k_0 and σ_0^2 twice while keeping α_0 unchanged. These three models were part of the five models used in Pritsker (1998). Note that the three models have the same marginal distribution $N(\alpha_0, V_E)$, where $V_E = \frac{\sigma^2}{2\kappa} = 0.01226$ is the same. Despite the stationary distribution being the same, the models offer different levels of dependence as quantified by the mean-reverting parameter κ . From Models -2 to 2, the process becomes more dependent as κ gets smaller. This is carefully designed by Pritsker to allow the effects of dependence to be investigated without changing the marginal density.

The region S was chosen based on the underlying transitional density so that the region attained more than 90% of the probability. In particular, for Models -2, 0 and 2, it was chosen by rotating respectively $[0.035, 0.25] \times [-0.03, 0.03]$, $[0.03, 0.22] \times [-0.02, 0.02]$ and $[0.02, 0.22] \times [-0.009, 0.009]$ 45 degrees anticlock-wise. The weight function $\omega(x, y) = |S|^{-1}I\{(x, y) \in S\}$ where $|S|$ is the area of S .

Both the cross-validation (CV) (Silverman 1986) and the reference to a bivariate normal distribution (the Scott Rule, Scott 1992) methods were used to guide the

selection of the bandwidth set \mathcal{H} whose values are reported in Table 1. Table 1 also contains the average bandwidths obtained by the two methods. We observed that, for each given n , regardless of which method was used, the chosen bandwidth became smaller as the model was shifted from Model -2 to Model 2. This indicated that both methods took into account the changing level of dependence induced by these models. The maximum likelihood estimator was used to estimate θ in each simulation and each resample in the bootstrap. Table 2 summarizes the quality of the parameter estimation.

The average sizes of the proposed test at the nominal size of 5% are reported in Table 3. It shows that the sizes of the proposed test were quite close to the nominal level for Vasicek Model -2 and Model 0 consistently for the three sample sizes considered. For Model 2, which has the weakest mean-reversion, there was some size distortion when $n = 125$. However, it was significantly alleviated when n was increased. The message conveyed by Table 3 is that we need not have a large number of years of data in order to achieve a reasonable size for the test. Table 3 also reports the single-bandwidth based test based on $N(h)$ and the asymptotic normality as conveyed by Theorem 1 with $J = 1$. Like Pritsker's study, this asymptotic test has severe size distortion too and highlights the need for implementing the bootstrap procedure.

The size distortion for Model 2 at $n = 125$ was partly due to the poor quality of parameter estimates. It is well known that the estimation of κ is subject to severe bias when the mean-reversion is weak. This is indeed confirmed in Table 2. The deterioration in the quality of the estimates, especially for κ when the dependence became stronger, is alarming. The relative bias of the κ -estimates was more than 200% for Model 2 at $n = 125$. It is nice to see that the proposed test did a very good job in producing a size of 12.6% under the severe circumstances. This indicates that

the proposed test is robust against poor parameter estimates.

We also carried out simulation for the test of Hong and Li (2005). The Scott rule adopted by Hong and Li was used to get an initial bandwidth (h_{scott}). We then chose 4 equally spaced bandwidths below and above the average h_{scott} . The nominal 5% test at each bandwidth was carried out with the lag values 1 and 4. We only report the results for lag value 1 in Table 4 as those for the other case were the same. For sample size not larger than 500, the sizes of the test did not settle well at the nominal level as reflected by the sizes being quite high for smaller bandwidths and then dropped quite quickly for larger bandwidths. The performance may be due to a combination of (i) the varying quality in parameter estimation as reported in Table 2 may affect the nature of the transformed series and (ii) the slow convergence to normality of the test statistic.

6.1.2 CIR Models

We then conduct simulation on three CIR models to see if the pattern of results observed for the Vasicek models holds for the CIR models. The CIR models are

$$dX_t = \kappa(\alpha - X_t)dt + \sigma\sqrt{X_t}dB_t, \quad (6.1)$$

where the parameters were: $\kappa = 0.89218$, $\alpha = 0.09045$ and $\sigma^2 = 0.032742$ in the first model (CIR 0); $\kappa = 0.44609$, $\alpha = 0.09045$ and $\sigma^2 = 0.16371$ in model CIR 1 and $\kappa = 0.22305$, $\alpha = 0.09045$ and $\sigma^2 = 0.08186$ in model CIR 2. CIR 0 was the model used in Prisker (1998) for power evaluation, which we used for power study as well. The three models mirror the Vasicek models 0, 1 and 2 of Pritsker (1998).

The region S was chosen by rotating 45-degrees anti-clockwise $[0.015, 0.25] \times [-0.015, 0.015]$ for CIR 0, $[0.015, 0.25] \times [-0.012, 0.012]$ for CIR 1 and $[0.015, 0.25] \times [-0.008, 0.008]$ for CIR 2 respectively. All the regions have a coverage probability of at least 0.90.

Table 5 reports the sizes of the proposed EL test as well as the single bandwidth based tests based on the bootstrap and the asymptotic normality. The bandwidth

sets were chosen based on the same principle as outlined for the Vasicek models and are reported in the table. The parameter estimation under these CIR models has the same pattern of quality as the Vasicek model as reported in Table 2. We find that the proposed test continued to have reasonable size for the three CIR models despite that there were severe bias in the estimation of κ . The size of the single bandwidth based tests as well as the combined test were quite respectable for a sample size of 125. We continue to see poor performance for the asymptotic test. It is interesting to see that despite the quality of the parameter estimation still being poor for the CIR 2, the severe size distortion observed earlier for Vasicek 2 was not present. Overall, the sizes for the CIR models were better than the corresponding Vasicek models, which seemed to suggest that the extra volatility offered by the CIR models “lubricates” the test performance. Hong and Li’s test was also performed for the three CIR models and are reported in Table 7. The performance was similar to that of the Vasicek models.

6.2. Power evaluation

To gain information on the power of the proposed test, we carried out simulation that tested for the Vasicek model while the real process was the CIR 0 as in Pritsker’s power evaluation of Aït-Sahalia’s test:

$$dX_t = \kappa(\alpha - X_t)dt + \sigma\sqrt{X_t}dB_t, \quad (6.2)$$

where $\kappa = 0.89218$, $\alpha = 0.09045$ and $\sigma^2 = 0.032742$. The region S was obtained by rotating $[0.015, 0.25] \times [-0.015, 0.015]$ 45 degrees anti-clock wise. The average CV bandwidths based on 500 simulations were 0.0202 (the standard error of 0.0045) for $n = 125$, 0.016991 (0.00278) for $n = 250$ and 0.014651 (0.00203) for $n = 500$.

Table 7 reports the power of the EL test and the single bandwidth based tests, as well as the bandwidth sets used in the simulation. We find the tests had quite good power. As expected, the power increased as n increased. One striking feature was that

the power of the test tends to be larger than the maximum of the single bandwidth based tests. This indicates that it possess attractive power properties. Table 7 also reports the power of the Hong and Li's test. It is found that the proposed test had better power for all the sample sizes considered.

7. Case studies. We apply the proposed test on the Federal fund rates between January 1963 and December 1998 which has $n = 432$ observations. Aït-Sahalia (1999) uses this data set to demonstrate the performance of the maximum likelihood estimation. We test for five popular one-factor diffusion models which have been proposed to model the dynamics of interest rates:

$$dX_t = \kappa(\alpha - X_t)dt + \sigma dB_t, \quad (7.1)$$

$$dX_t = \kappa(\alpha - X_t)dt + \sigma\sqrt{X_t}dB_t, \quad (7.2)$$

$$dX_t = X_t\{\kappa - (\sigma^2 - \kappa\alpha)X_t\}dt + \sigma X_t^{3/2}dB_t, \quad (7.3)$$

$$dX_t = \kappa(\alpha - X_t)dt + \sigma X_t^\rho dB_t, \quad (7.4)$$

$$dX_t = (\alpha_{-1}X_t^{-1} + \alpha_0 + \alpha_1X_t + \alpha_2X_t^2)dt + \sigma X_t^{3/2}dB_t. \quad (7.5)$$

They are respectively the Ornstein-Uhlenbeck process (7.1) proposed by Vasicek (1977), the CIR model (7.2), the inverse of the CIR process (7.3), the constant elasticity of volatility (CEV) model (7.4) and the nonlinear drift model (7.5) of Aït-Sahalia (1996).

The data are displayed in Figure 1, which indicates a strong dependence as they scattered around a narrow band around the 45-degree line. There was an increased volatility when the rate was larger than 12%. The model-implied transitional densities are displayed in Figure 2 using the MLEs given in Aït-Sahalia (1999), which were used in the formulation of the proposed test statistic. Figure 2 shows that the densities implied by the Inverse CIR, the CEV and the nonlinear drift models were similar to each other, and were quite different from those of the Vasicek and CIR models.

The bandwidths prescribed by the Scott rule and the CV for the kernel estimation were respectively $h_{\text{ref}} = 0.007616$ and $h_{\text{cv}} = 0.00129$. Plotting the density surfaces indicated that a reasonable range for h was from 0.007 to 0.02, which offered a range of smoothness from slightly undersmoothing to slightly oversmoothing. This led to a bandwidth set consisting of 7 bandwidths with $h_{\text{min}} = 0.007$, $h_{\text{max}} = 0.020$ and $a = 0.8434$.

Kernel transitional density estimates and the smoothed model-implied transitional densities for the five models are plotted in Figure 3 for $h = 0.007$. By comparing Figure 2 with Figure 3, we notice the effect of kernel smoothing on these model-implied densities.

In formulating the final test statistic L_n , we chose

$$N(h) = \frac{1}{n} \sum_{t=1}^n \ell\{\tilde{p}_{\hat{\theta}}(X_{t+1}|X_t)\} \omega_1(X_t, X_{t+1}), \quad (7.6)$$

where ω_1 is a uniform weight over a region by rotating $[0.005, 0.4] \times [-0.03, 0.03]$ 45 degree anticlock-wise. The region contains all the data pairs (X_t, X_{t+1}) . As seen from (7.6), $N(h)$ is asymptotically equivalent to the statistic defined in (3.8) with $\omega(x, y) = p(x, y)\omega_1(x, y)$, and has the same flavor with the test statistic of Aït-Sahalia (1996).

The p-values of the proposed tests as well as the tests based on single bandwidth for each of the five diffusion models are reported in Table 8, which were obtained based on 500 bootstrap resamples. It shows little empirical support for the Vasicek model and quite weak support for the CIR. What is surprising is that there was quite some empirical support for the inverse CIR, the CEV and the nonlinear drift models. In particular, for CEV and the nonlinear drift models, the p-values of the single bandwidth based tests were all quite supportive even for small bandwidths. Indeed, by looking at Figure 3, we see quite noticeable agreement between the nonparametric kernel density estimates and the smoothed densities implied by the CEV and nonlinear

drift models.

8. Conclusion. The proposed test differs from the test of Hong and Li (2005) in three aspects. The proposed test is based on a direct comparison between the kernel estimate and the smoothed model-implied transitional density, whereas Hong and Li's test is an indirect comparison via data transformation. An advantage of the direct approach is its robustness against poor quality parameter estimation which is often the case for weak mean-reverting diffusion models. This is because both the shape and the orientation of the transitional density are much less affected by the poor quality parameter estimation. The second aspect is that Hong and Li's test is based on asymptotic normality and can be under the influence of slow convergence discussed earlier although the transformed series is asymptotically independent. The last aspect is that our proposed test is based on a set of bandwidths, which makes the test robust against the choice of bandwidths.

The conclusion we draw from our studies is that the kernel method works effectively for testing of diffusion models and is capable of modeling dependence induced by a diffusion model. It is clear from the studies of Pritsker (1998) and this paper that a proper implementation is vitally important. After all, the kernel method is just an instrument for constructing nonparametric curve estimates. For a complex task of testing a diffusion model, it will not work automatically by itself and requires other procedures to make it work. However, what is working is the idea of comparing the kernel estimate of a characteristic curve and the corresponding model-implied curve of a diffusion model. This is the main idea of Aït-Sahalia (1996). The role of the kernel method is in translating the idea into some raw discrepancy measure. Anything beyond it, for instance, the test statistic formulation and the choice of the critical value should be the responsibilities of the other procedures.

9. Acknowledgments. The authors thank Cheng Yong Tang for excellent

computing assistance and a National University of Singapore Academic research grant for generous support. The second author acknowledges support of an Australian Research Council Discovery Grants Program.

Appendix. As the Lagrange multiplier $\lambda(x, y)$ is implicitly dependent on h , we need first to extend the convergence rate for a single h -based $\sup_{(x,y) \in S} \lambda(x, y)$ conveyed in (3.6) uniformly over the bandwidth set \mathcal{H} . To prove Theorem 1, we need the following lemmas first.

Lemma A.1. *Under Assumptions 1–4, $\max_{h \in \mathcal{H}} \sup_{(x,y) \in S} \lambda(x, y) = o_p\{n^{-1/4} \log(n)\}$.*

Proof: For any $\delta > 0$

$$P \left(\max_{h \in \mathcal{H}_n} \sup_{(x,y) \in S} h \lambda(x, y) \geq \delta n^{-1/2} \log(n) \right) \leq \sum_{h \in \mathcal{H}} P \left(\sup_{(x,y) \in S} h \lambda(x, y) \geq \delta n^{-1/2} \log(n) \right).$$

As the number of bandwidths in H is finite, by checking the relevant derivations in Chen, Härdle and Li (2002), it can be shown that

$$P \left(\sum_{(x,y) \in S} h \lambda(x, y) \geq \delta n^{-1/2} \log(n) \right) \rightarrow 0$$

as $n \rightarrow \infty$. This implies that $\max_{h \in \mathcal{H}} \sup_{(x,y) \in S} h \lambda(x, y) = o_p\{\delta n^{-1/2} \log(n)\}$. Then the lemma is established by noting that h_{\min} , the smallest bandwidth in \mathcal{H}_n , is of order $n^{-\gamma_1}$ where $\gamma_1 \in (0, 1/4)$ as assumed in Assumption 1. \square

Before introducing the next lemma, we present some expansions for the EL test statistic $N(h)$. Let

$$\tilde{p}_\theta(x, y) = \tilde{p}_\theta(y|x) \hat{\pi}(x) \quad \text{and} \quad \tilde{p}(x, y) = n^{-1} \sum_{t=1}^{n+1} K_h(x - X_t) \sum_{s=1}^{n+1} w_s(y) p(X_s|X_t)$$

be kernel smooths of the parametric and real underlying joint densities $p_\theta(x, y)$ and $p(x, y)$ respectively. Due to the relationship between transitional and joint densities, from (3.8)

$$\begin{aligned} N(h) &= (nh^2) \int \int \frac{\{\hat{p}(x, y) - \tilde{p}_\theta(x, y)\}^2}{R^2(K)p(x, y)} \omega(x, y) dx dy + \tilde{O}_p\{h^2 + (nh^2)^{-1/2} \log^3(n)\} \\ &= (nh^2) R^{-2}(K) \int \int \left[\frac{\{\hat{p}(x, y) - \tilde{p}(x, y)\}^2}{p(x, y)} + \frac{2\{\hat{p}(x, y) - \tilde{p}(x, y)\}\{\tilde{p}(x, y) - \tilde{p}_\theta(x, y)\}}{p(x, y)} \right] \end{aligned}$$

$$\begin{aligned}
& + \frac{\{\tilde{p}(x, y) - \tilde{p}_{\tilde{\theta}}(x, y)\}^2}{p(x, y)} \Big] \omega(x, y) dx dy + \tilde{O}_p\{h^2 + (nh^2)^{-1/2} \log^3(n)\} \\
& =: N_1(h) + N_{2\tilde{\theta}}(h) + N_{3\tilde{\theta}}(h) + \tilde{O}_p\{h^2 + (nh^2)^{-1/2} \log^3(n)\}
\end{aligned} \tag{A.1}$$

Here and throughout the proofs, $\tilde{o}(\delta_n)$ and $\tilde{O}(\delta_n)$ denote stochastic quantities which are respectively $o(\delta_n)$ and $O(\delta_n)$ uniformly over S for a non-negative sequence $\{\delta_n\}$.

Using Assumptions 3.3 and 3.4, we have $N_{l\tilde{\theta}}(h) = N_{l\theta^*}(h) + \tilde{O}_p(n^{-1/2})$ where $\theta^* = \theta_0$ under H_0 and θ_1 under H_1 . Thus,

$$N(h) = N_1(h) + N_{2\theta^*}(h) + N_{3\theta^*}(h) + \tilde{O}_p\{h^2 + (nh^2)^{-1/2} \log^3(n)\}$$

We start with some lemmas on $\hat{p}(x, y)$, $\tilde{p}(x, y)$ and $\tilde{p}_{\theta}(x, y)$. Let $K^{(2)}$ be the convolution of K , $MK^{(2)}(t) = \int \int uK(u)K(t+u)du$ and $p_3(x, y, z)$ be the joint density of (X_t, X_{t+1}, X_{t+2}) .

Lemma A.2. *Under Assumptions 1-4, we have,*

$$\begin{aligned}
\text{Cov}\{\hat{p}(s_1, t_1), \hat{p}(s_1, t_1)\} &= \frac{K^{(2)}\left(\frac{s_2-s_1}{h}\right) K^{(2)}\left(\frac{s_2-s_1}{h}\right) p(s_1, t_1)}{nh^2} \\
&- \frac{MK^{(2)}\left(\frac{s_2-s_1}{h}\right) \frac{\partial p(s_1, t_1)}{\partial x} + MK^{(2)}\left(\frac{t_2-t_1}{nh}\right) \frac{\partial p(s_1, t_1)}{\partial y}}{nh} \\
&+ \frac{p_3(s_1, t_1, t_2)K^{(2)}\left(\frac{s_2-t_1}{h}\right) + p_3(s_2, t_2, t_1)K^{(2)}\left(\frac{s_1-t_2}{h}\right)}{nh} \\
&+ o\{(nh)^{-1}\}.
\end{aligned}$$

Proof: Let $Z_t(s, t) = K_h(s - X_t)K_h(t - X_{t+1})$ so that $\hat{p}(s, t) = n^{-1} \sum_{t=1}^n Z_t(s, t)$ and

$$\begin{aligned}
\text{Cov}\{\hat{p}(s_1, t_1), \hat{p}(s_1, t_1)\} &= n^{-1} \text{Cov}\{Z_1(s_1, t_1), Z_1(s_2, t_2)\} \\
&+ \frac{1}{n-1} [\text{Cov}\{Z_1(s_1, t_1), Z_2(s_2, t_2)\} + \text{Cov}\{Z_2(s_1, t_1), Z_1(s_2, t_2)\}] \\
&+ Q_n,
\end{aligned} \tag{A.2}$$

where $Q_n = n^{-1} \sum_{l=2}^{n-1} (1-ln^{-1}) [\text{Cov}\{Z_1(s_1, t_1), Z_{l+1}(s_2, t_2)\} + \text{Cov}\{Z_{l+1}(s_1, t_1), Z_1(s_2, t_2)\}]$.

Standard derivations show that

$$\begin{aligned}
\text{Cov}\{Z_1(s_1, t_1), Z_1(s_2, t_2)\} &= \frac{K^{(2)}\left(\frac{s_2-s_1}{h}\right) K^{(2)}\left(\frac{s_2-s_1}{h}\right) p(s_1, t_1)}{h^2} \\
&- \frac{MK^{(2)}\left(\frac{s_2-s_1}{h}\right) \frac{\partial p(s_1, t_1)}{\partial x} + MK^{(2)}\left(\frac{t_2-t_1}{h}\right) \frac{\partial p(s_1, t_1)}{\partial y}}{nh} + \tilde{O}(1) \\
\text{Cov}\{Z_1(s_1, t_1), Z_2(s_2, t_2)\} &= \frac{p_3(s_1, t_1, t_2)K^{(2)}\left(\frac{s_2-t_1}{h}\right)}{h} + \tilde{O}(1)
\end{aligned} \tag{A.3}$$

Apply Davydov inequality for α -mixing sequences via the same route of Fan and Gijbels (1996, P.251) that $Q_n = \tilde{o}\{(nh)^{-1}\}$. This together with (A.2) - (A.3) lead prove the lemma.

□

That α -mixing leads to $Q_n = \tilde{o}\{(nh)^{-1}\}$, the pre-whitening effect of smoothing bandwidth, is a fact that we will use repeatedly without further mentioning.

Lemma A.3. *Suppose that Assumptions 1–4 hold. Let $\Delta_\theta(x, y) = \{p_\theta(y|x) - p(y|x)\}\pi(x)$. Then*

$$E\{\tilde{p}_\theta(x, y) - \hat{p}(x, y)\} = \Delta_\theta(x, y) + \frac{1}{2}h^2\sigma_K^2\left\{\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right\}\Delta_\theta(x, y) + \tilde{O}(h^3), \quad (\text{A.4})$$

$$E\{\tilde{p}_\theta(x, y) - \tilde{p}(x, y)\} = \Delta_\theta(x, y) + \frac{1}{2}h^2\sigma_K^2\left\{\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right\}\Delta_\theta(x, y) + \tilde{O}(h^3), \quad (\text{A.5})$$

$$\text{Cov}\{\tilde{p}(s_1, t_1), \tilde{p}(s_2, t_2)\} = \frac{K^{(2)}\left(\frac{t_2 - t_1}{h}\right)p(s_1, t_1)p(s_2, t_1)}{nh\pi(t_2)} + \tilde{o}\{(nh)^{-1}\}. \quad (\text{A.6})$$

Proof: As $\tilde{p}_\theta(x, y) = n^{-1} \sum_{t=1}^{n+1} K_h(x - X_t) \sum_{s=1}^{n+1} w_s(y) p_\theta(X_s|X_t)$, from the bias of the local linear regression (Fan and Gijbels, 1996) and kernel density estimators, and that the transitional density has uniformly bounded third derivatives in Assumption 3,

$$\begin{aligned} E\{\tilde{p}_\theta(x, y)\} &= E\left\{n^{-1} \sum_{t=1}^{n+1} K_h(x - X_t) \left\{p_\theta(y|X_t) + \frac{1}{2}h^2\sigma_K^2 \frac{\partial^2 p_\theta(y|X_t)}{\partial y^2} + \tilde{O}_p(h^3)\right\}\right\} \\ &= p_\theta(y|x)\pi(x) + \frac{1}{2}h^2\sigma_K^2\left\{\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right\}p_\theta(y|x)\pi(x) + \tilde{O}(h^3). \end{aligned}$$

Then employing the same kind of derivation on $E\{\hat{p}(x, y)\}$, we readily establish (A.4).

For the purpose of deriving (A.6), we note the following expansion for $\tilde{p}(x, y)$ based on the notion of equivalent kernel for local linear estimator (Fan and Gijbels, 1996):

$$\begin{aligned} \tilde{p}(x, y) &= n^{-1} \sum K_h(x - X_t) \sum_{s=1}^{n+1} w_s(y) p_\theta(X_s|X_t) \\ &= n^{-1} \sum K_h(x - X_t) \sum_{s=1}^{n+1} \pi^{-1}(y) K_h(y - X_s) p_\theta(X_s|X_t) \{1 + \tilde{o}_p(h)\} \\ &= \pi(x)\pi^{-1}(y) \sum_{s=1}^{n+1} \pi^{-1}(y) K_h(y - X_s) p_\theta(X_s|x) \{1 + \tilde{o}_p(h)\}. \end{aligned} \quad (\text{A.7})$$

Then derivations similar to those used in Lemma A.2 readily establish (A.6). □

Lemma A.4. *Under Assumptions 1–4, we have*

$$\text{Cov}\{\hat{p}(s_1, t_1), \tilde{p}(s_2, t_2)\} = \frac{p(s_1, t_1)}{nh\pi(t_2)} \left[K^{(2)}\left(\frac{t_2 - s_1}{h}\right) p(s_2, s_1) + K^{(2)}\left(\frac{t_2 - s_1}{h}\right) p(s_2, s_1) \right]$$

Proof: From (A.7),

$$\begin{aligned}
& \text{Cov}\{\hat{p}(s_1, t_1), \tilde{p}(s_2, t_2)\} \\
&= \frac{\pi(s_2)}{n\pi(t_2)} \left[\text{Cov}\{K_h(s_1 - X_1)K_h(t_1 - X_2), K_h(t_2 - X_1)p(X_1|s_2)\} \right. \\
&+ \left. \text{Cov}\{K_h(s_1 - X_1)K_h(t_1 - X_2), K_h(t_2 - X_2)p(X_2|s_2)\} \right] \{1 + \tilde{o}(1)\} \\
&= \frac{\pi(s_2)}{nh\pi(t_2)} \left\{ K^{(2)} \left(\frac{t_2 - s_1}{h} \right) p(s_1|s_2)p(s_1, t_1) + K^{(2)} \left(\frac{t_2 - t_1}{h} \right) p(t_1|s_2)p(s_1, t_1) \right\} \{1 + \tilde{o}(1)\}
\end{aligned}$$

which then leads to the statement of the lemma. \square

Lemma A.5. *If H_0 is true, then $N_{2\theta^*}(h) = N_{3\theta^*}(h) = 0$ for all $h \in \mathcal{H}$.*

Proof: Under H_0 , $p(y|x) = p_{\theta_0}(y|x)$ and $\theta^* = \theta_0$. Hence, $\tilde{p}(x, y) - \tilde{p}_{\theta^*}(x, y) = n^{-1} \sum K_h(x - X_t) \sum w_s(y) \{p(X_s|X_t) - p_{\theta_0}(X_s|X_t)\} = 0$. This completes the proof. \square

Lemma A.6. *Suppose that Assumptions 1–4 hold. If H_1 is true, then*

$$N_{2\theta^*}(h) = \tilde{O}_p\{(nh^2)^{-1/2} \log(n)\} \times \int \int \frac{\Delta_{\theta_1}(x, y)}{p(x, y)} \omega(x, y) dx dy \quad \text{and} \quad (\text{A.8})$$

$$N_{3\theta^*}(h) = (nh^2)R^{-2}(K) \int \int \frac{\Delta^2(x, y)}{p(x, y)} \omega(x, y) dx dy \{1 + \tilde{o}_p(1)\} \quad (\text{A.9})$$

uniformly over the bandwidth set \mathcal{H} .

Proof: Under H_1 , $\theta^* = \theta_1$. Standard derivation as those in Lemma A.3 show that

$$\begin{aligned}
\tilde{p}(x, y) - \tilde{p}_{\theta_1}(x, y) &= \tilde{p}(x, y) - E\{\tilde{p}(x, y)\} + E\{\tilde{p}(x, y)\} - E\{\tilde{p}_{\theta_1}(x, y)\} \\
&+ E\{\tilde{p}_{\theta_1}(x, y)\} - \tilde{p}_{\theta_1}(x, y) \\
&= -\Delta_{\theta_1}(x, y) + \tilde{O}_p\{(nh)^{-1/2} \log(n) + h^2\}.
\end{aligned} \quad (\text{A.10})$$

This and the fact that $\hat{p}(x, y) - \tilde{p}(x, y) = \tilde{O}\{(nh^2)^{-1/2} \log(n) + h^3\} = \tilde{O}\{(nh^2)^{-1/2} \log(n)\}$ lead to (A.8). And (A.9) can be argued similarly using (A.10). \square

Let us now study the leading term $N_1(h)$. From (A.1) and by hiding the variables of integrations,

$$\begin{aligned}
N_1(h) &= \frac{(nh^2)}{R^2(K)} \int \int \frac{\{\hat{p} - \tilde{p}\}^2}{R^2(K)p} \omega \\
&= \frac{(nh^2)}{R^2(K)} \int \int \left[\frac{\{\hat{p} - E\hat{p}\}^2}{p} + \frac{\{E\tilde{p} - \tilde{p}\}^2}{p} + \frac{\{E\hat{p} - E\tilde{p}\}^2}{p} \right] \omega
\end{aligned}$$

$$\begin{aligned}
& + \frac{2\{\hat{p} - E\hat{p}\}\{E\hat{p} - E\tilde{p}\}}{p} + \frac{2\{\hat{p} - E\hat{p}\}\{E\tilde{p} - \tilde{p}\}}{p} + \frac{2\{E\hat{p} - E\tilde{p}\}\{E\tilde{p} - \tilde{p}\}}{p} \Big] \omega \\
& =: \sum_{j=1}^6 N_{1j}(h)
\end{aligned} \tag{A.11}$$

We are to show in the following lemmas that $N_{11}(h)$ dominates $N_1(h)$ and $N_{1j}(h)$ for $j \geq 2$ are all negligible except $N_{12}(h)$ which contributes to the mean of $N_1(h)$ in the second order.

Lemma A.7. *Under Assumptions 1–4, then uniformly with respect to \mathcal{H}*

$$h^{-1}E\{N_{11}(h) - 1\} = o(1), \tag{A.12}$$

$$\text{Var}\{h^{-1}N_{11}(h)\} = \frac{2K^{(4)}(0)}{R^4(K)} \int \int \omega^2(x, y) dx dy + o(1), \tag{A.13}$$

$$\text{Cov}\{h_1^{-1}N_{11}(h_1), h_2^{-1}N_{11}(h_2)\} = \frac{2\nu(h_1/h_2)}{R^4(K)} \int \int \omega^2(x, y) dx dy + o(1). \tag{A.14}$$

Proof: From Lemma A.1 and note that $MK^{(2)}(0) = 0$

$$\begin{aligned}
E\{N_{11}(h)\} &= \frac{nh^2}{R^2(K)} \int \int \frac{\text{Var}\{\hat{p}(x, y)\}}{p(x, y)} \omega(x, y) dx dy \\
&= \frac{1}{R^2(K)} \int \int \left[\left(K^{(2)}(0) \right)^2 + 2hK^{(2)} \left(\frac{y-x}{h} \right) \frac{p_3(x, y, y)}{p(x, y)} \right] \\
&\times \omega(x, y) dx dy \{1 + o(1)\} = 1 + O(h^2),
\end{aligned} \tag{A.15}$$

which leads to (A.12). To derive (A.13), let

$$\hat{Z}_n(s, t) = (nh^2)^{1/2} \frac{\hat{p}(s, t) - E\hat{p}(s, t)}{R(K)p^{1/2}(x, t)}.$$

It may be shown from the fact that K is bounded and other regularity condition assumed that $E\{|\hat{Z}_n(s_1, t_1)|^{2+\epsilon} |\hat{Z}_n(s_2, t_2)|^{2+\epsilon}\} \leq M$ for some positive ϵ and M . And hence $\{\hat{Z}_n(s, t)\}_{n \geq 1}$ and $\{\hat{Z}_n^2(s_1, t_1)\}_{n \geq 1}$ are uniformly integrable respectively. Also, $\left(\hat{Z}_n(s_1, t_1), \hat{Z}_n(s_2, t_2) \right)^T \xrightarrow{d} (Z(s_1, t_1), Z(s_2, t_2))^T$ which is a bivariate normal random variable with mean zero and a covariance matrix

$$\Sigma = \begin{pmatrix} 1 & g\{(s_1, t_1), (s_2, t_2)\} \\ g\{(s_1, t_1), (s_2, t_2)\} & 1 \end{pmatrix}$$

where $g\{(s_1, t_1), (s_2, t_2)\} = K^{(2)}\left(\frac{s_2-s_1}{h}\right) K^{(2)}\left(\frac{t_2-t_1}{h}\right) \frac{p^{1/2}(s_1, t_1)}{R(K)p^{1/2}(s_2, t_2)}$. Hence,

$$\begin{aligned}
\text{Var}\{N_{11}(h)\} &= \int \int \int \int \text{Cov}\{\hat{Z}_n^2(s_1, s_2), \hat{Z}_n^2(s_2, t_2)\} \omega(s_1, t_1) \omega(s_2, t_2) ds_1 dt_1 ds_2 dt_2 \\
&= \int \int \int \int \text{Cov}\{Z^2(s_1, s_2), Z^2(s_2, t_2)\} \omega(s_1, t_1) \omega(s_2, t_2) ds_1 dt_1 ds_2 dt_2 \\
&= 2 \int \int \int \int \text{Cov}^2\{Z(s_1, s_2), Z(s_2, t_2)\} \omega(s_1, t_1) \omega(s_2, t_2) ds_1 dt_1 ds_2 dt_2 \\
&= \frac{2}{R^4(K)} \int \int \int \int \left\{ K^{(2)}\left(\frac{s_2-s_1}{h}\right) K^{(2)}\left(\frac{t_2-t_1}{h}\right) \right\}^2 \frac{p(s_1, t_1)}{R^2(K)p(s_2, t_2)} \\
&\quad \times \omega(s_1, t_1) \omega(s_2, t_2) ds_1 dt_1 ds_2 dt_2 \\
&= \frac{2h^2 K^{(4)}(0)}{R^4(K)} \int \int \omega^2(s, t) \pi^{-2}(t) ds dt
\end{aligned} \tag{A.16}$$

the second to fourth equations are valid up to a factor $\{1 + o(1)\}$. In the third equation above, we use a fact regarding the fourth product moments of normal random variables. Combining (A.15) and (A.16), (A.12) and (A.13) are derived. It is trivial to check that it is valid uniformly for all $h \in \mathcal{H}$.

The proof for (A.14) follows that for (A.13). \square

Lemma A.8. *Under Assumptions 1–4, then uniformly with respect to $h \in \mathcal{H}$*

$$h^{-1} N_{12}(h) = \frac{1}{R(K)} \int \int \frac{p(x, y)}{\pi(y)} \omega(x, y) dx dy + o_p(1), \tag{A.17}$$

$$h^{-1} N_{1j}(h) = o_p(1) \quad \text{for } j \geq 3 \tag{A.18}$$

Proof: From (A.6) and the fact that $K^{(2)}(0) = R(K)$,

$$\begin{aligned}
E\{N_{12}(h)\} &= \frac{(nh^2)}{R^2(K)} \int \int \frac{\text{Var}\{\tilde{p}\}}{p} \omega \\
&= \frac{h}{R(K)} \int \int \frac{p(x, y)}{\pi(y)} \omega(x, y) dx dy \{1 + \tilde{o}(1)\}.
\end{aligned} \tag{A.19}$$

Let $\tilde{Z}_n(s, t) = (nh)^{1/2} \frac{\tilde{p}(s, t) - E\tilde{p}(s, t)}{R^{1/2}(K)p(x, t)}$. Using the same approach that derives (A.16) and from Lemma A.2

$$\begin{aligned}
\text{Var}\{N_{12}(h)\} &= \frac{h^2}{R^2(K)} \int \int \int \int \text{Cov}\{\tilde{Z}_n^2(s_1, s_2), \tilde{Z}_n^2(s_2, t_2)\} \omega(s_1, t_1) \omega(s_2, t_2) ds_1 dt_1 ds_2 dt_2 \\
&= \frac{2h^2}{R^4(K)} \int \int \int \int \left\{ K^{(2)}\left(\frac{t_2-t_1}{h}\right) \right\}^2 \pi^{-2}(t_2) \omega(s_2, t_2) ds_1 dt_1 ds_2 dt_2 \\
&= \frac{2h^4 K^{(4)}(0)}{R^4(K)} \int \int \omega^2(s, t) \pi^{-2}(t) ds dt
\end{aligned} \tag{A.20}$$

up to a factor $\{1 + \tilde{o}(1)\}$. Combining (A.19) and (A.20), (A.17) is derived. It can be checked that it is valid uniformly for all $h \in \mathcal{H}$.

As $\{E\hat{p}(x, y) - E\tilde{p}(x, y)\}^2 = \tilde{O}(h^6)$ and $h = o(n^{-1/7})$ for all $h \in \mathcal{H}$, we have $(nh^2)h^6 = o_p(h)$ uniformly for all $h \in \mathcal{H}$ and hence (A.15) for $j = 2$.

Obviously $E\{N_{14}(h)\} = 0$. Use the same method for $\text{Var}\{N_{12}(h)\}$ and from Lemmas A.1 and A.2,

$$\begin{aligned} \text{Var}\{N_{14}(h)\} &= \frac{4(nh^2)^2}{R^4(K)} \int \int \int \int \frac{E\{\hat{p} - \tilde{p}\}(s_1, t_1) E\{\hat{p} - \tilde{p}\}(s_2, t_2) \text{Cov}\{\hat{p}(s_1, t_1), \hat{p}(s_2, t_2)\}}{p(s_1, t_1)p(s_2, t_2)} \\ &\quad \times w(s_1, t_1)w(s_2, t_2)ds_1dt_1ds_2dt_2 \end{aligned}$$

As $E\{\hat{p} - \tilde{p}\}(s_1, t_1)E\{\hat{p} - \tilde{p}\}(s_2, t_2) = O(h^6)$ and the integral over the covariance produces a h^2 in addition to $(nh^2)^{-1}$. Thus, $\text{Var}\{N_{14}(h)\} = O\{(nh^2)h^6h^2\} = o(h^2)$ uniformly for all $h \in \mathcal{H}$. This means that $N_{14} = o_p(h)$ uniformly for all $h \in \mathcal{H}$. Using exactly the same derivation, but employing Lemma A.3 instead of Lemma A.1, we have $N_{16} = o_p(h)$ uniformly for all $h \in \mathcal{H}$.

It remains to study $N_{15}(h)$. From Lemma A.3,

$$\begin{aligned} E\{N_{15}(h)\} &= -\frac{2(nh^2)}{R^2(K)} \int \int \frac{\text{Cov}\{\hat{p}(x, y), \tilde{p}(x, y)\}}{p(x, y)} \omega(x, y) dx dy \\ &= -\frac{4h}{R^2(K)} \int \int \frac{K^{(2)}\left(\frac{y-x}{h}\right) p(x, y)}{\pi(y)} \omega(x, y) dx dy \\ &= -\frac{4h^2}{R^2(K)} \int K^{(2)}(u) du \int \pi(y) \omega(y, y) dy \{1 + \tilde{o}(1)\}. \end{aligned}$$

using Assumption 1(iv). It may be shown using the same method that derives (A.20) that $\text{Var}\{N_{15}(h)\} = o(h^2)$ and hence $N_{15}(h) = o_p(h)$.

Let $L(h) = \frac{1}{C(K)h}\{N(h) - 1\}$ and $\beta = \frac{1}{C(K)R(K)} \int \int \frac{p(x, y)}{\pi(y)} \omega(x, y) dx dy$. In view of Lemmas A.5, A.7 and A.8, we have under H_0 , uniformly with respect to \mathcal{H} ,

$$L(h) = \frac{1}{C(K)h}\{N_{11}(h) - 1\} + \beta + o_p(1) \quad (\text{A.21})$$

Define $L_1(h) = \frac{1}{C(K)h}\{N_{11}(h) - 1\}$.

Lemma A.9. *Under Assumptions 1–4 and H_0 , as $n \rightarrow \infty$,*

$$(L_1(h_1), \dots, L_1(h_J)) \xrightarrow{d} N_J(0, \Sigma_J).$$

Proof: According to the Cramér–Wold device, in order to prove Lemma A.9, it suffices to show that

$$\sum_{i=1}^k c_i L_1(h_i) \xrightarrow{d} N_J(0, c^\tau \Sigma_J c) \quad (\text{A.22})$$

for an arbitrary vector of constants $c = (c_1, \dots, c_k)^\tau$. Without loss of generality, we will consider only the proof for the case of $k = 2$. To apply Lemma A.1 of Gao and King (2003), we introduce the following notation. For $i = 1, 2$, define $d_i = \frac{c_i}{\sqrt{2}h_i}$ and $\xi_t = (X_t, X_{t+1})$,

$$\begin{aligned} \epsilon_{ti}(x, y) &= K\left(\frac{x - X_t}{h_i}\right) K\left(\frac{y - X_{t+1}}{h_i}\right) - E\left[K\left(\frac{x - X_t}{h_i}\right) K\left(\frac{y - X_{t+1}}{h_i}\right)\right], \\ \phi_i(\xi_s, \xi_t) &= \frac{1}{nh_i^2} \int \int \frac{\epsilon_{si}(x, y) \epsilon_{ti}(x, y)}{p(x, y) R^2(K)} \omega(x, y) dx dy, \\ \phi_{st} &= \phi(\xi_s, \xi_t) = \sum_{i=1}^2 d_i \phi_i(\xi_s, \xi_t) \quad \text{and} \quad \bar{L}_1(h_1, h_2) = \sum_{t=2}^T \sum_{s=1}^{t-1} \phi_{st}. \end{aligned} \quad (\text{A.23})$$

It is noted that for any given $s, t \geq 1$ and fixed x and y , $E[\phi(x, \xi_t)] = E[\phi(\xi_s, y)] = 0$. Since the asymptotic variance $c^\tau \Sigma_J c$ is a non-random quadratic form depending neither on h nor on n , in order to apply their Lemma A.1, it suffices to verify

$$\max\{M_n, N_n\} h_1^{-2} \rightarrow 0 \quad \text{as } n \rightarrow \infty, \quad (\text{A.24})$$

where

$$\begin{aligned} M_n &= \max \left\{ n^2 M_{n1}^{\frac{1}{1+\delta}}, n^2 M_{n51}^{\frac{1}{2(1+\delta)}}, n^2 M_{n52}^{\frac{1}{2(1+\delta)}}, n^2 M_{n6}^{\frac{1}{2}} \right\} \\ N_n &= \max \left\{ n^{\frac{3}{2}} M_{n21}^{\frac{1}{2(1+\delta)}}, n^{\frac{3}{2}} M_{n22}^{\frac{1}{2(1+\delta)}}, n^{\frac{3}{2}} M_{n3}^{\frac{1}{2}}, n^{\frac{3}{2}} M_{n4}^{\frac{1}{2(1+\delta)}}, n^{\frac{3}{2}} M_{n7}^{\frac{1}{1+\delta}} \right\}, \end{aligned}$$

in which

$$\begin{aligned} M_{n1} &= \max_{1 \leq i < j < k \leq n} \max \left\{ E|\phi_{ik}\psi_{jk}|^{1+\delta}, \int |\phi_{ik}\theta_{jk}|^{1+\delta} dP(\xi_i) dP(\xi_j, \xi_k) \right\}, \\ M_{n21} &= \max_{1 \leq i < j < k \leq n} \max \left\{ E|\phi_{ik}\phi_{jk}|^{2(1+\delta)}, \int |\phi_{ik}\phi_{jk}|^{2(1+\delta)} dP(\xi_i) dP(\xi_j, \xi_k) \right\}, \\ M_{n22} &= \max_{1 \leq i < j < k \leq n} \max \left\{ \int |\phi_{ik}\phi_{jk}|^{2(1+\delta)} dP(\xi_i, \xi_j) dP(\xi_k), \int |\phi_{ik}\phi_{jk}|^{2(1+\delta)} dP(\xi_i) dP(\xi_j) dP(\xi_k) \right\}, \\ M_{n3} &= \max_{1 \leq i < j < k \leq n} E|\phi_{ik}\phi_{jk}|^2, \quad M_{n4} = \max_{\substack{1 < i, j, k \leq 2n \\ i, j, k \text{ different}}} \left\{ \max_P \int |\phi_{1i}\phi_{jk}|^{2(1+\delta)} dP \right\}, \end{aligned}$$

where the maximization over P in the equation for M_{n4} is taken over the four probability measures $P(\xi_1, \xi_i, \xi_j, \xi_k)$, $P(\xi_1)P(\xi_i, \xi_j, \xi_k)$, $P(\xi_1)P(\xi_{i_1})P(\xi_{i_2}, \xi_{i_3})$, and $P(\xi_1)P(\xi_i)P(\xi_j)P(\xi_k)$, where (i_1, i_2, i_3) is the permutation of (i, j, k) in ascending order;

$$\begin{aligned} M_{n51} &= \max_{1 \leq i < j < k \leq n} \max \left\{ E \left| \int \phi_{ik} \phi_{jk} \phi_{ik} \phi_{jk} dP(\xi_i) \right|^{2(1+\delta)} \right\}, \\ M_{n52} &= \max_{1 \leq i < j < k \leq n} \max \left\{ \int \left| \int \phi_{ik} \phi_{jk} \phi_{ik} \phi_{jk} dP(\xi_i) \right|^{2(1+\delta)} dP(\xi_j) dP(\xi_k) \right\}, \\ M_{n6} &= \max_{1 \leq i < j < k \leq n} E \left| \int \phi_{ik} \phi_{jk} dP(\xi_i) \right|^2, \quad M_{n7} = \max_{1 \leq i < j < n} E \left[|\phi_{ij}|^{1+\delta} \right]. \end{aligned}$$

Due to the fact that ϕ_{st} is only a linear combination of $\phi_1(X_s, X_t)$ and $\phi_2(X_s, X_t)$, in order to verify the above conditions, it suffices to verify that each $\phi_i(X_s, X_t)$ satisfies the conditions. In the following, we will only deal with the case of $i = 1$, as the other case follows similarly.

Without any confusion, we replace h_1 by h for simplicity. To verify the M_n part of (A.24), we verify only

$$\lim_{n \rightarrow \infty} n^2 h^{-2} M_{n1}^{\frac{1}{1+\delta}} = 0. \quad (\text{A.25})$$

Let $q(x, y) = \omega(x, y)p^{-1}(x, y)$ and

$$\psi_{ij} = \frac{1}{nh^2} \int K((x - X_i)/h) K((y - X_{i+1})/h) K((x - X_j)/h) K((y - X_{j+1})/h) q(x, y) dx dy$$

for $1 \leq i < j < k \leq n$. Direct calculation implies

$$\begin{aligned} \psi_{ik} \psi_{jk} &= (nh^2)^{-2} \int \cdots \int K\left(\frac{x - X_i}{h}\right) K\left(\frac{y - X_{i+1}}{h}\right) K\left(\frac{x - X_k}{h}\right) K\left(\frac{y - X_{k+1}}{h}\right) q(x, y) \\ &\quad \times K\left(\frac{u - X_j}{h}\right) K\left(\frac{v - X_{j+1}}{h}\right) K\left(\frac{u - X_k}{h}\right) K\left(\frac{v - X_{k+1}}{h}\right) q(u, v) dx dy du dv \\ &= b_{ijk} + \delta_{ijk}, \end{aligned}$$

where $\delta_{ijk} = \psi_{ik} \psi_{jk} - b_{ijk}$ and

$$\begin{aligned} b_{ijk} &= n^{-2} q(X_i, X_{i+1}) q(X_j, X_{j+1}) \\ &\quad \times K^{(2)}\left(\frac{X_i - X_k}{h}\right) K^{(2)}\left(\frac{X_j - X_k}{h}\right) K^{(2)}\left(\frac{X_{i+1} - X_{k+1}}{h}\right) K^{(2)}\left(\frac{X_{j+1} - X_{k+1}}{h}\right), \end{aligned}$$

in which $K^{(2)}(x) = \int K(y) K(x + y) dy$.

For any given $1 < \zeta < 2$ and n sufficiently large, we may show that

$$\begin{aligned}
M_{n11} &= E \left[|\psi_{ij}\psi_{ik}|^\zeta \right] \leq 2 \left(E \left[|b_{ijk}|^\zeta \right] + E \left[|\delta_{ijk}|^\zeta \right] \right) = 2E \left[|b_{ijk}|^\zeta \right] (1 + o(1)) \\
&= n^{-2\zeta} \int \int \int |q(x, y)q(u, v)|^\zeta \left| K^{(2)} \left(\frac{x-z}{h} \right) K^{(2)} \left(\frac{u-z}{h} \right) \right|^\zeta \\
&\quad \times \left| K^{(2)} \left(\frac{y-w}{h} \right) K^{(2)} \left(\frac{v-w}{h} \right) \right|^\zeta p(x, y, u, v, z, w) dx dy du dv dz dw \\
&= C_1 n^{-2\zeta} h^4
\end{aligned} \tag{A.26}$$

where $p(x, y, u, v, z, w)$ denotes the joint density of $(X_i, X_{i+1}, X_j, X_{j+1}, X_k, X_{k+1})$ and C_1 is a constant. Thus, as $n \rightarrow \infty$

$$n^2 h^{-2} M_{n11}^{\frac{1}{1+\delta}} = C n^2 h^{-1} \left(n^{-2\zeta} h^2 \right)^{1/\zeta} = h^{2\frac{(2-\zeta)}{\zeta}} \rightarrow 0. \tag{A.27}$$

Hence, (A.27) shows that (A.25) holds for the first part of M_{n1} . The proof for the second part of M_{n1} follows similarly. As for (A.26), we have that as $n \rightarrow \infty$

$$\begin{aligned}
M_{n3} &= E |\psi_{ik}\psi_{jk}|^2 = (nh^2)^{-4} h^8 \int \int \int |q(x, y)q(u, v)|^2 \left| K^{(2)} \left(\frac{x-z}{h} \right) K^{(2)} \left(\frac{y-z}{h} \right) \right|^2 \\
&\quad \times \left| K^{(2)} \left(\frac{u-w}{h} \right) K^{(2)} \left(\frac{v-w}{h} \right) \right|^2 p(x, y, z, u, v, w) dx dy dz du dv dw \\
&= C_1 n^{-4} h^4,
\end{aligned}$$

where C_2 is a constant. This implies that as $n \rightarrow \infty$

$$n^{3/2} h^{-2} M_{n3}^{\frac{1}{2}} = C n^{-1/2} \rightarrow 0. \tag{A.28}$$

Thus, (A.28) now shows that (A.24) holds for M_{n3} . It follows from the structure of $\{\psi_{ij}\}$ that (A.24) holds automatically for M_{n51} , M_{n52} and M_{n6} . We now start to prove that (A.25) holds for M_{n21} . For some $0 < \delta < 1$ and $1 \leq i < j < k \leq n$, let $M_{n21} = E [|\psi_{ik}\psi_{jk}|^{2(1+\delta)}]$. Similarly to (A.26) and (A.27), we obtain that as $n \rightarrow \infty$

$$n^{3/2} h^{-2} M_{n21}^{\frac{1}{2(1+\delta)}} \rightarrow 0. \tag{A.29}$$

This finally completes the proof of (A.25) for M_{n21} and thus (A.25) holds for the first part of $\{\phi_1(X_s, X_t)\}$. Similarly, one can show that (A.24) holds for the other parts of $\{\phi_1(X_s, X_t)\}$. Thus, we have shown that equation (A.22) holds for the case of $k = 2$ under H_0 . \square

Proof of Theorem 1: From (A.21) and Lemma A.9, we have under H_0

$$(L(h_1), \dots, L(h_J)) \xrightarrow{d} N_J(\beta \mathbf{1}_J, \Sigma_J).$$

Let $Z = (Z_1, \dots, Z_k)^T \stackrel{d}{\sim} N_J(\beta \mathbf{1}_k, \Sigma_J)$. By the mapping theorem, under H_0 ,

$$L_n = \max_{h \in \mathcal{H}} L(h) \xrightarrow{d} \max_{1 \leq k \leq J} Z_k. \quad (\text{A.30})$$

Hence the theorem is established. \square .

Let $l_{0\alpha}$ be the upper- α quantile of $\max_{1 \leq i \leq k} Z_i$. As the distribution of $N_J(\beta \mathbf{1}_k, \Sigma_J)$ is free of n , so is that of $\max_{1 \leq i \leq k} Z_i$. And hence $l_{0\alpha}$ is a fixed quantity with respect to n .

The following lemmas are required for the proof of Theorem 2.

Lemma A.10. *Under Assumptions 1–4 and H_1 , for any fixed real value x , as $n \rightarrow \infty$,*

$$P(L_n \geq x) \rightarrow 1.$$

Proof: Let $A = R^{-2}(K) \int \int \frac{\Delta^2(x, y)}{p(x, y)} \omega(x, y) dx dy$. From Lemmas A.6 and A.7, under H_1 , $(nh^2)^{-1}N(h) = A + o_p(1)$ uniformly with respect to \mathcal{H} . Hence, $(nh)^{-1}L(h) = \frac{A}{C(K)} + o_p(1)$ for all $h \in \mathcal{H}$. Hence, for some $i \in \{1, \dots, k\}$, $P(L_n < x) \leq P\{L(h_i) < x\} = P\{(nh)^{-1}L(h_i) < (nh)^{-1}x\}$. As $(nh)^{-1}L(h_i) \xrightarrow{p} \frac{A}{C(K)} > 0$ and $(nh)^{-1}x \rightarrow 0$, hence $P(L_n < x) \rightarrow 0$. \square

We now turn to the bootstrap EL test statistic $N^*(h)$, which is a version of $N(h)$ based on $\{X_t^*\}_{t=1}^{n+1}$ generated according to the parametric transitional density $p_{\tilde{\theta}}$. Let $\hat{p}(x, y)$ and $\tilde{p}_{\tilde{\theta}}^*(x, y)$ are the bootstrap versions of $\hat{p}(x, y)$ and $\tilde{p}(x, y)$ respectively, and $\tilde{\theta}^*$ be the maximum likelihood estimate based the bootstrap sample. Then, the following analogue expansion to (A.2) is valid for $N^*(h)$

$$\begin{aligned} N^*(h) &= (nh^2) \int \int \frac{\{\hat{p}^*(x, y) - \tilde{p}_{\tilde{\theta}^*}^*(x, y)\}^2}{R^2(K)p_{\tilde{\theta}}(x, y)} \omega(x, y) dx dy + \tilde{o}_p\{h\} \\ &= N_1^*(h) + N_{2\tilde{\theta}}^*(h) + N_{3\tilde{\theta}}^*(h) + \tilde{o}_p(h) \end{aligned}$$

where $N_j^*(h)$ for $j = 1, 2$ and 3 are the bootstrap version of $N_j(h)$ respectively. As the bootstrap resample is generated according to $p_{\tilde{\theta}}$, the same arguments which lead to Lemma 5 mean that $N_2^*(h) = N_3^*(h) = 0$. Thus, $N^*(h) = N_1^*(h) + \tilde{o}_p(1)$ where

$$N_1^*(h) = (nh^2) \int \int \frac{\{\hat{p}^*(x, y) - \tilde{p}^*(x, y)\}^2}{R^2(K)p_{\tilde{\theta}}(x, y)} \omega(x, y) dx dy$$

And similar lemmas to Lemmas A.7 and A.8 can be established to study $N_{1j}^*(h)$ which are the bootstrap version of $N_{1j}^*(h)$ respectively.

Let $L^*(h) = \frac{1}{C(K)h} \{N(h) - 1\}$ and $\hat{\beta} = \frac{1}{C(K)R(K)} \int \int \frac{p_{\hat{\theta}}(x,y)}{\pi_{\hat{\theta}}(y)} \omega(x,y) dx dy$.

Lemma A.11. Under Assumptions 1–4, as $n \rightarrow \infty$, given $\{X_t\}_{t=1}^{n+1}$,

$$(L^*(h_1), \dots, L^*(h_k)) \xrightarrow{d} N_J(\hat{\beta}1_J, \Sigma_J).$$

Proof: The proof follows that of Theorem 1 with the real underlying transitional density $p_{\hat{\theta}}$ to replace p in the proof of Theorem 1. \square

Let l_{α}^* be the upper- α conditional quantile of $L_n^* = \max_{h \in \mathcal{H}} L^*(h)$ given $\{X_t\}_{t=1}^{n+1}$.

Proof of Theorem 2: (i) Let $\hat{Z} = (\hat{Z}_1, \dots, \hat{Z}_k)^T$ such that its conditional distribution given $\{X_t\}_{t=1}^{n+1}$ is $N_J(\hat{\beta}1_J, \Sigma_J)$ and $l_{0\alpha}^*$ be the upper- α conditional quantile of $\max_{1 \leq i \leq k} \hat{Z}_i$. As $\tilde{\theta} = \theta + O_p(n^{-1/2})$ as assumed in Assumption 5 (i), $\hat{\beta} = \beta + O_p(n^{-1/2})$. This means that $\max_{1 \leq i \leq k} \hat{Z}_i = \max_{1 \leq i \leq k} Z_i + o_p(1)$. Therefore, $l_{0\alpha}^* = l_{0\alpha} + o_p(1)$. From Lemma A.11, we have via the mapping theorem again, $\lim_{n \rightarrow \infty} P\{L_n^* \geq l_{0\alpha}^* | \{X_t\}_{t=1}^{n+1}\} = \alpha$. Therefore,

$$l_{\alpha}^* = l_{0\alpha} + o_p(1). \quad (\text{A.31})$$

As $L_n \xrightarrow{d} \max_{1 \leq i \leq k} Z_i$, by Slutsky theorem,

$$P(L_n \geq l_{\alpha}^*) = P(L_n + o_p(1) \geq l_{0\alpha}) \rightarrow P(\max_{1 \leq i \leq k} Z_i \geq l_{0\alpha}) = \alpha$$

which completes the part (i) of Theorem 2. The part (ii) of Theorem 2 is a direct consequence of Lemma A.9 and (A.31). \square

REFERENCES

- Aït-Sahalia, Y. (1996). Testing continuous-time models of the spot interest rate. *Review of Financial Studies* 9 385–426.
- Aït-Sahalia, Y. (1999). Transition densities for interest rate and other nonlinear diffusions. *Journal of Finance* 54 1361–1395.
- Aït-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. *Econometrica* 70 223–262.
- Aït-Sahalia, Y., Bickel, P. and Stoker, T. (2001). Goodness-of-fit tests for regression using kernel methods. *Journal of Econometrics* 105 363–412.

- Bandi, F. and Phillips, P. C. B. (2003). Fully nonparametric estimation of scalar diffusion models. *Econometrica* **71** 241–284.
- Brown, B. and Chen, S. X. (1998). Combined and least squares empirical likelihood. *Annals of the Institute of Statistical Mathematics* **50** 697–714.
- Cai, Z. and Hong, Y. (2003). Nonparametric methods in continuous-time finance: A selective review, In *Recent Advances and Trends in Nonparametric Statistics* (M.G. Akritas and D.M. Politis, eds.), 283–302.
- Chapman, D. A. and Pearson, N. D. (2000). Is the short rate drift actually nonlinear. *Journal of Finance* **55** 355–388.
- Chen, S. X. and Cui, H. J. (2004). On the second order properties of empirical likelihood for generalized estimation equations. Technical Report, Department of Statistics, Iowa State University.
- Chen, S. X., Härdle, W. and Li, M. (2002). An empirical likelihood Goodness-of-fit test for time series. Technical Report 2002-2, National University of Singapore, <http://www.stat.nus.edu.sg/>.
- Chen, S. X., Härdle, W. and Li, M. (2003). An empirical likelihood goodness-of-fit test for time series. *Journal of the Royal Statistical Society Series B* **65** 663–678.
- Cox, J. C., Ingersoll, J. E. and Ross, S. A. (1985). A theory of term structure of interest rates. *Econometrica* **53** 385–407.
- Fan, J. (1996). Test of significance based on wavelet thresholding and Neyman’s truncation. *Journal of the American Statistical Association* **434** 674–688.
- Fan, J. (2005). A selective overview of nonparametric methods in financial econometrics. *Statistical Science* to appear.
- Fan, J. and Gijbels, I. (1996), *Local Polynomial Modeling and Its Applications*. Chapman and Hall, London.
- Fan, J. and Yao, Q. (2003) *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer-Verlag, New York.
- Fan, J., Yao, Q. and Tong, H. (1996). Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. *Biometrika* **83** 189–196.
- Fan, J. and Yim, T. H. (2004). A data-driven method for estimating conditional densities. *Biometrika* **91** 819–834.
- Fan, J. and Zhang, C. (2003). A Re-examination of Stanton’s diffusion estimation with applications to financial model validation. *Journal of the American Statistical Association* **457** 118–134.
- Fan, J. and Zhang, J. (2004) Sieve empirical likelihood ratio tests for nonparametric functions. *Annals of Statistics* **32** 1858–1907.
- Fan, J., Zhang, C. and Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks phe-

- nomenon. *Annals of Statistics* **29** 153–193.
- Fan, Y. and Li, Q. (1996). Consistent model specification tests: omitted variables and semiparametric functional forms. *Econometrica* **64** 865–890.
- Gao, J. and King, M. L. (2003). Estimation and model specification testing in nonparametric and semiparametric regression models. Unpublished report, The University of Western Australia. Available at www.maths.uwa.edu.au/~jiti/jems.pdf.
- Genon-Caralot, V., Jeantheau, T. and Laredo, C. (2000). Stochastic volatility models as hidden Markov models and statistical applications. *Bernoulli* **6** 1051–1079.
- Gozalo, P. L. and Linton, O. (2001). Testing additivity in generalized nonparametric regression models with estimated parameters. *Journal of Econometrics* **104** 1–48.
- Härdle, W. (1990). *Applied Nonparametric Regression*. Cambridge University Press, Cambridge.
- Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Annals of Statistics* **21** 1926–1947.
- Hart, J. (1997). *Nonparametric Smoothing and Lack-of-Fit Tests*. Springer, New York.
- Hjellvik, V. and Tjøstheim, D. (1995). Nonparametric tests of linearity for time series. *Biometrika* **82** 351–368.
- Hjellvik, V., Yao, Q. and Tjøstheim, D. (1998). Linearity testing using local polynomial approximation. *Journal of Statistical Planning and Inference* **68** 295–321.
- Hjort, N. L., McKeague, I. W. and Van Keilegom, I. (2005). Extending the scope of empirical likelihood. Under revision for *Annals of Statistics*.
- Hong, Y. and Li, H. (2005). Nonparametric specification testing for continuous-time models with application to spot interest rates. *Review of Financial Studies* **18** 37–84.
- Horowitz, J. L. and Spokoiny, V. G. (2001). An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica* **69** 599–632.
- Hyndman, R. J. and Yao, Q. (2002). Nonparametric estimation and symmetry tests for conditional density functions. *Journal of Nonparametric Statistics* **14** 259–278.
- Jiang, G. and Knight, J. (1997). A nonparametric approach to the estimation of diffusion processes with an application to a short-term interest rate model. *Econometric Theory* **13** 615–645.
- Kreiss, J., Neumann, M. H. and Yao, Q. (1998). Bootstrap tests for simple structures in nonparametric time series regression. Discussion Paper, Humboldt-Universität Zu Berlin.
- Kitamura, Y. (1997). Empirical likelihood methods with weakly dependent processes, *Annals of Statistics* **25** 2084–2102.
- Li, G. (2003). Nonparametric likelihood ratio goodness-of-fit tests for survival data. *Journal of Multivariate Analysis* **86** 166–182.

- Li, G. and Van Keilegom, I. (2002). Likelihood ratio confidence bands in non-parametric regression with censored data. *Scandinavian Journal of Statistics* **29** 547–562.
- Li, Q. (1999). Consistent model specification tests for time series econometric models. *Journal of Econometrics* **92** 101–147.
- Müller, H.-G. (1991). Smooth optimum kernel estimators near endpoints, *Biometrika* **78** 521–530.
- Müller, H.-G. (1993). On the boundary kernel method for non-parametric curve estimation near endpoints. *Scandinavian Journal of Statistics* **20** 313–328.
- Müller, H. G. and Stadtmüller, U. (1999). Multivariate boundary kernels and a continuous least squares principle. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **61** 439–458.
- Owen, A. (1988). Empirical likelihood ratio confidence regions for a single functional. *Biometrika* **75** 237–249.
- Owen, A. (1990). Empirical Likelihood Ratio Confidence Regions. *Annals of Statistics* **18** 90–120.
- Pritsker, M. (1998). Nonparametric density estimation and tests of continuous time interest rate models. *Review of Financial Studies* **11** 449–487.
- Qin, J. and J. Lawless (1994). Empirical Likelihood and General Estimating Functions, *Annals of Statistics* **22** 300–325.
- Robinson, P. (1989). Hypothesis testing in semiparametric and nonparametric models for econometric time series. *Review of Economic Studies* **56** 511–534.
- Scott, D. W. (1992). *Multivariate Density Estimation*. John Wiley, New York.
- Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.
- Stanton, R. (1997). A nonparametric model of term structure dynamics and the market price of interest rate risk. *Journal of Finance* **52** 1973–2002.
- Tripathi, G. and Kitamura, Y. (2003) Testing conditional moment restrictions. *Annals of Statistics* **31** 2059–2095.
- Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics* **5** 177–188.
- Yu, J. and Phillips, P. (2001). A Gaussian approach for continuous time models of the short-term interest rate. *Econometrics Journal* **4** 210–224.

Table 1: Smoothing bandwidths in the simulation of the Three Vasicek Models

	Model -2		Model 0		Model 2	
n	h_{cv}	h_{ref}	h_{cv}	h_{ref}	h_{cv}	h_{ref}
$n = 125$	0.0319	0.0276	0.0181	0.0161	0.0095	0.0093
	(0.005)	(0.0017)	(0.003)	(0.0014)	(0.0017)	(0.0007)
$n = 250$	0.0261	0.0247	0.0163	0.0146	0.0090	0.0080
	(0.004)	(0.0015)	(0.002)	(0.0012)	(0.0014)	(0.001)
500	0.0231	0.0220	0.0140	0.013	0.0080	0.0073
	(0.003)	(0.0013)	(0.0017)	(0.0007)	(0.001)	(0.0008)

Table 2: Relative Bias (RBIAS) and Variance of the MLES for the Vasicek Models.

Model -2	$\kappa = 3.334$		$\alpha = 0.0891$		$\sigma^2 = 0.093498$	
n	RBIAS	VAR	RBIAS	VAR	RBIAS	VAR
120	13.78%	1.134	0.088%	6.591e-5	1.004%	5.056e-5
250	5.19%	0.5132	0.382%	3.477e-5	0.393%	2.456e-5
500	3.987%	0.24714	0.057%	1.685e-5	0.1%	9.912e-6
Model 0	$\kappa = 0.429$		$\alpha = 0.0891$		$\sigma^2 = 0.033057$	
120	54.33%	0.347	0.218%	2.612e-4	1%	1.035e-5
250	21.18%	0.125	0.722%	1.396e-4	0.449%	4.911e-6
500	12.31%	0.054	0.102%	6.669e-5	0.008%	2.016e-6
Model 2	$\kappa = 0.215$		$\alpha = 0.0891$		$\sigma^2 = 0.023375$	
120	238.35%	0.240	3.792%	0.0103	0.898%	2.549e-6
250	96.05%	0.056	1.274%	6.640e-4	0.485%	1.134e-6
500	50.96%	0.022	0.122%	2.682e-4	0.001%	4.793e-7

Table 3: Empirical Size (in percentage) of the Combined EL Test and the Single Bandwidth Based Test (in the middle) for the Vasicek Models, as well as that of the Single Bandwidth Test Based on the Asymptotic Normality (in round bracket).
A: Model -2

	Bandwidth						Adaptive Test
$n = 125$	0.03	0.0320	0.0340	0.0362	0.0386	0.0411	
Size	9.4 (40.4)	8.2 (38.8)	5.2 (34.8)	4.6 (34.8)	3 (34.2)	2.4 (34.8)	4.4
$n = 250$	0.022	0.0231	0.0243	0.0256	0.0269	0.0284	
Size	8 (34.4)	5.2 (28.6)	4.6 (24)	4.4 (21)	3.4 (17.4)	2.6 (16)	4.6
$n = 500$	0.02	0.0210	0.0221	0.0233	0.0245	0.0258	
Size	6.2 (29.6)	5.8 (23.6)	5.4 (19)	5.2 (14.6)	5 (10.8)	5 (8.4)	5.4

B: Model 0

	Bandwidths						Adaptive Test
$n = 125$	0.016	0.0173	0.0187	0.0203	0.022	0.0237	
Size	5.8 (43)	6 (39)	6 (36.6)	4.2 (34.8)	4.4 (36.6)	3 (37)	4.2
$n = 250$	0.014	0.0153	0.0167	0.0182	0.0198	0.0216	
Size	6 (31.6)	6.2 (27)	6.2 (20.6)	3.8 (20)	2.4 (17.8)	2.8 (17.8)	5.2
$n = 500$	0.0106	0.0114	0.0124	0.0134	0.0145	0.0157	
Size	6.8 (36.4)	4.4 (26.8)	5.2 (20.6)	6.4 (13)	5.6 (11.2)	4 (9.2)	5.4

C: Model 2

	Bandwidths						Adaptive Test
$n = 125$	0.008	0.009	0.0101	0.0114	0.01281	0.01441	
Size	12.6 (60)	11 (53.4)	10 (47.2)	14.6 (46)	14.4 (45)	13.6 (42.2)	12.6
$n = 250$	0.006	0.0067	0.0076	0.0085	0.0095	0.0107	
Size	12.2 (39)	10 (35)	7.4 (31)	8.8 (30)	7 (31)	11 (33.2)	8.8
$n = 500$	0.004	0.0047	0.0054	0.0063	0.0074	0.0086	
Size	8.2 (75.2)	8.4 (63)	8 (51.6)	8.6 (39.8)	7 (32.8)	9 (24.4)	7.2

Table 4: Empirical Size (in percentage) of Hong and Li's Test for Vasicek Models.

	Bandwidths									h_{scott}
$n = 125$	0.08	0.09	0.1	0.11	0.12	0.13	0.14	0.15	0.16	0.1301 (0.0026)
Vasicek -2	11.8	6.2	3.8	2.4	1.2	0.6	0.6	0.6	0.6	
Vasicek 0	13.2	7.8	4.2	2	1.2	1.2	0.8	0.8	0.8	
Vasicek 2	11.8	7	4.2	2.4	2	1.6	1.4	1.6	1.6	
$n = 250$	0.07	0.08	0.09	0.1	0.11	0.12	0.13	0.14	0.15	0.1156 (0.0017)
Vasicek -2	15	9.6	6.2	3.8	2.2	1.6	1.6	1.2	1	
Vasicek 0	13	8.6	5.8	4.4	3.4	2.6	1.8	1.8	2	
Vasicek 2	15	10.6	7.4	4.2	3	2.6	1.6	1.4	1.4	
$n = 500$	0.06	0.07	0.08	0.09	0.1	0.11	0.12	0.13	0.14	0.1026 (0.0010)
Vasicek -2	15.6	12.4	7.6	4.4	2.8	2.2	2.4	1.6	1.2	
Vasicek 0	19.4	13	8.4	5.4	3.4	2.6	2.6	2.4	1.6	
Vasicek 2	17	14	9	6.2	4.6	3.8	3.2	2.8	2.4	
$n = 1000$	0.05	0.06	0.07	0.08	0.09	0.1	0.11	0.12	0.13	0.09138 (0.00065)
Vasicek -2	22.6	17.2	13.2	7	4.8	3.8	2.4	2.2	1.8	
Vasicek 0	22.6	17.8	13.8	8.6	4.8	4	3	3	2.2	
Vasicek 2	21.2	16.4	12.8	8	5.2	3.8	3.2	2.8	2.6	

Table 5: Empirical Size (in percentage) of the Combined EL Test and the Single Bandwidth Based Test (in the middle) for the CIR Models, as well as that of the Single Bandwidth Test Based on the Asymptotic Normality (in round bracket).
A: CIR 0

	Bandwidths						Adaptive Test
$n = 125$	0.022	0.0252	0.0289	0.0332	0.0380	0.0436	
Size	4.8 (39.6)	4.8 (38.8)	4.8 (33.6)	3.2 (35.2)	2.4 (34.8)	2 (38.4)	3.0
$n = 250$	0.018	0.0207	0.0238	0.0275	0.0317	0.0365	
Size	5.2 (33)	5.6 (31.4)	5.2 (27.4)	5.2 (23.4)	4 (24.4)	3.8 (28.4)	5.0
$n = 500$	0.016	0.0182	0.0207	0.0236	0.0269	0.0307	
Size	4.2 (31.4)	5.4 (23.8)	4.6 (21)	4.8 (17.6)	3.6 (16)	4.2 (14.6)	4.8

B: CIR 2

	Bandwidths						Adaptive Test
$n = 125$	0.017	0.0196	0.0226	0.0261	0.03	0.0346	
Size	5.4 (68.4)	3.6 (65.6)	4.0 (63.2)	4.2 (62.6)	3.2 (60.8)	2.6 (52)	3.8
$n = 250$	0.014	0.016	0.0184	0.0211	0.0243	0.0279	
Size	5.2 (54.8)	6.6 (38.2)	5.6 (34.2)	5.6 (32.4)	5.6 (33.6)	3.4 (30.4)	5.2
$n = 500$	0.012	0.0138	0.0159	0.0183	0.0211	0.0243	
Size	5.4 (42)	3.8 (35.2)	4.4 (29)	4.4 (26.6)	4.6 (22.8)	4 (19.8)	5.2

C: CIR 3

	Bandwidths						Adaptive Test
$n = 125$	0.012	0.0138	0.0158	0.0182	0.0209	0.024	
Size	7.6 (67)	7.2 (61.8)	7.6 (51.8)	6.2 (41.8)	6.6 (41.8)	4.2 (58.4)	6.8
$n = 250$	0.01	0.0115	0.0132	0.0152	0.0174	0.02	
Size	5.6 (59.2)	6.4 (54)	5.8 (50)	6.8 (44.4)	6.2 (43.6)	5.4 (48.6)	6.2
$n = 500$	0.008	0.0092	0.0105	0.0121	0.0139	0.016	
Size	4 (34.4)	4.2 (31.4)	3.6 (25.6)	4 (25.8)	4.8 (22.4)	3.4 (23.6)	4

Table 6: Empirical Size (in percentage) of Hong and Li's Test for CIR Models

	Bandwidths									h_{scott} (SE)
$n = 125$	0.08	0.09	0.1	0.11	0.12	0.13	0.14	0.15	0.16	
CIR -2	9.4	5.8	4.6	3.2	2	1.4	0.6	0.8	0.8	0.129(0.0027)
CIR 0	10.8	8	5.6	3.4	2.8	2.6	1.8	1.6	1.6	0.1288(0.0063)
CIR 2	10	8	6	4.4	4	3.8	3.4	2.6	1.8	0.1272 (0.0153)
$n = 250$	0.07	0.08	0.09	0.1	0.11	0.12	0.13	0.14	0.15	
CIR -2	12.2	7.8	5.2	3.6	2.4	1.6	1.8	1.8	1.8	0.1148(0.00172)
CIR 0	12.6	9	5.8	4	2.4	2.4	1.8	1.4	1.2	0.1149(0.00168)
CIR 2	15.6	11	7.4	5.2	3.6	3.2	2.8	2.6	2.6	0.1144(0.0090)
$n = 500$	0.06	0.07	0.08	0.09	0.1	0.11	0.12	0.13	0.14	
CIR -2	18.2	15.4	10	5.2	3.2	3	2.4	2.2	2	0.1023(0.00108)
CIR 0	20	15.6	9.6	5.4	4.2	3.4	2.8	2.4	2	0.1024(0.00104)
CIR 2	19.6	14.4	9.8	5	4.6	4	3.6	2.6	2.2	0.1025(0.001018)
$n = 1000$	0.05	0.06	0.07	0.08	0.09	0.1	0.11	0.12	0.13	0.09138 (0.00065)
CIR -2	21	14.6	11.8	8.6	7	4.2	2.4	2	1.8	0.0910(0.000624)
CIR 0	20.2	14.8	12	10	7.6	4.8	2.6	2.6	2.8	0.0911(0.000606)
CIR 2	21.4	16	10.2	7.8	7.2	5.6	3.8	2.8	2.4	0.0912(0.00060)

Table 7**A: Power (in percentage) of the Adaptive EL Test**

n		Single Bandwidth Based Tests					Adaptive Test
$n = 125$	h	0.0199	0.0219	0.0241	0.0265	0.0291	
	Power	80.4	74	67.2	66.4	65.2	79.8
$n = 250$	h	0.0141	0.0158	0.0177	0.0199	0.0223	
	Power	87.6	81.2	76.4	74	72.8	88.6
$n = 500$	h	0.0113	0.0126	0.0141	0.0157	0.0175	
	Power	90.8	84.8	82.8	84.4	80.8	96.8

B: Power (in percentage) of Hong and Li's Test

n		Single Bandwidth Based Tests									$h_{\text{scott}}(\text{SE})$
$n = 125$	h	0.08	0.09	0.1	0.11	0.12	0.13	0.14	0.15	0.16	0.1283(0.003)
	Power	26	21.6	16.4	12.6	9.6	6.6	5.8	5.6	4.8	
$n = 250$	h	0.07	0.08	0.09	0.1	0.11	0.12	0.13	0.14	0.15	0.1135(0.0020)
	Power	41	34	29	24.2	18.8	16.6	14.6	12.2	10.4	
$n = 500$	h	0.06	0.07	0.08	0.09	0.1	0.11	0.12	0.13	0.14	0.1007(0.0013)
	Power	57.4	54	49	45	40.2	36	34	32.8	31.2	

Table 8: P-values for the Federal Fund Rate Data

Model		Single Bandwidth Based Tests							Adaptive
		0.007	0.0083	0.010	0.012	0.014	0.017	0.020	
Vasicek	Test Stats	29.71	23.64	16.81	9.64	3.52	-0.47	-1.18	29.71
	$l_{0.05}^*$	-2.49	-2.80	-3.04	-3.45	-3.79	-3.64	-2.19	2.54
	P-Values	0.00	0.00	0.00	0.00	0.00	0.00	0.016	0.00
CIR	Test Stats	3.40	0.32	-1.74	-2.19	-0.34	4.09	12.80	12.80
	$l_{0.05}^*$	-3.12	-3.66	-3.89	-3.52	-0.45	6.63	22.27	22.27
	P-Values	0.00	0.00	0.00	0.014	0.044	0.09	0.142	0.142
ICIR	Test Stats	5.80	9.13	14.06	21.97	32.72	46.58	66.63	66.63
	$l_{0.05}^*$	-1.62	2.68	10.84	30.19	66.83	145.5	303.4	303.4
	P-Values	0.002	0.018	0.044	0.078	0.146	0.216	0.294	0.294
CEV	Test Stats	5.06	8.16	12.86	20.56	31.11	44.76	64.65	64.65
	$l_{0.05}^*$	3.57	14.78	14.78	37.98	88.75	185.3	434.7	434.77
	P-Values	0.032	0.10	0.064	0.136	0.238	0.33	0.434	0.434
NL	Test Stats	5.87	9.21	14.22	22.45	33.69	48.15	69.10	69.10
	$l_{0.05}^*$	1.00	9.77	27.47	66.06	128.5	281.2	557.5	557.52
	P-Values	0.018	0.054	0.11	0.172	0.266	0.344	0.422	0.422

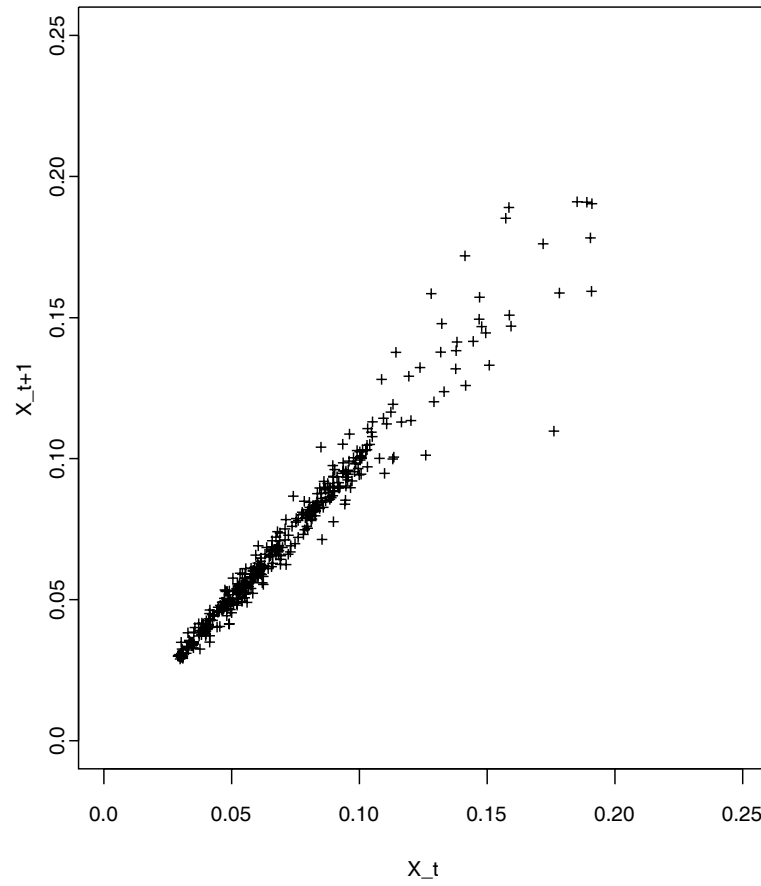


Figure 1. Scatter Plot of the Federal Fund Rate Data.

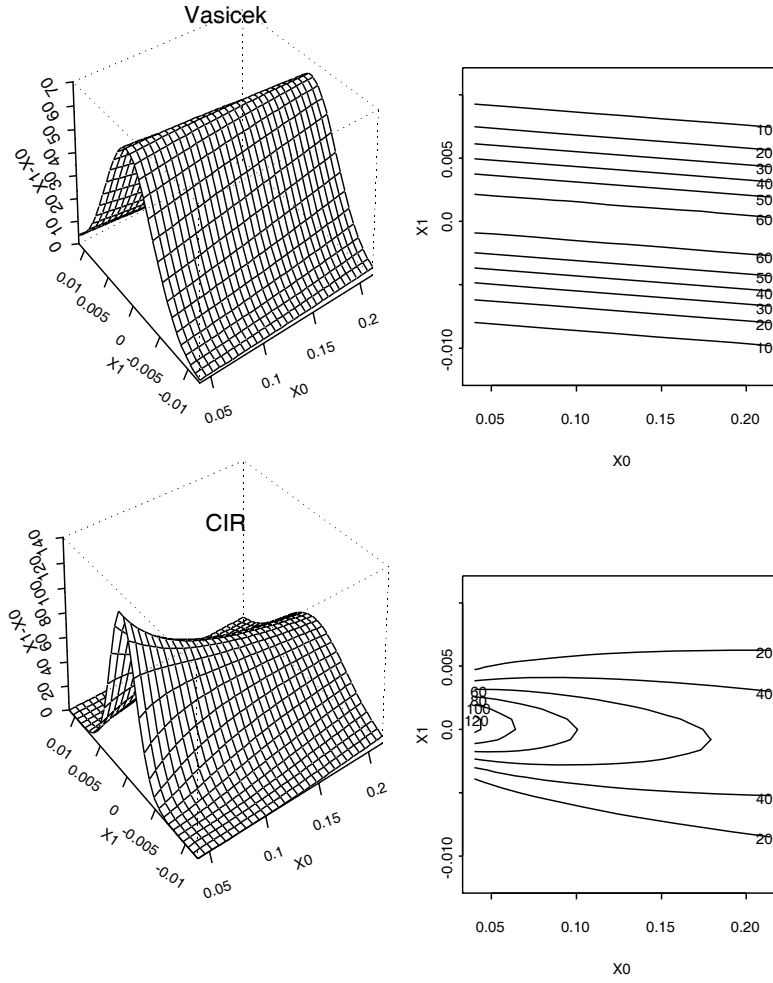


Figure 2 (a). Transitional densities of the Vasicek and CIR models after rotating 45 degree clock-wise. The right panels are the contours of the density surface explicated in the right panels.

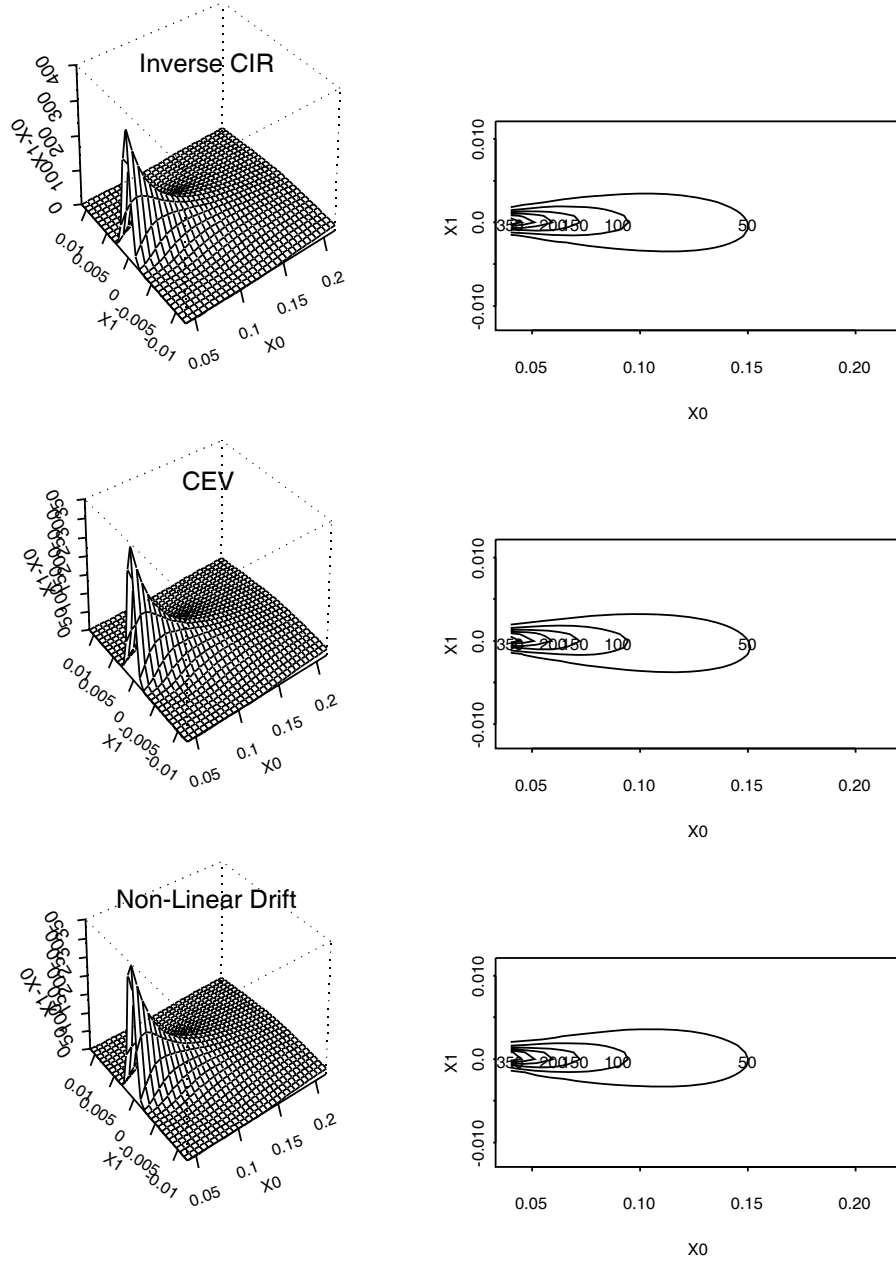


Figure 2 (b). Transitional densities of the inverse CIR, the CEV and the nonlinear drift models after rotating 45 degree clock-wise. The right panels are the contours of the density surface exhibited in the left panels.

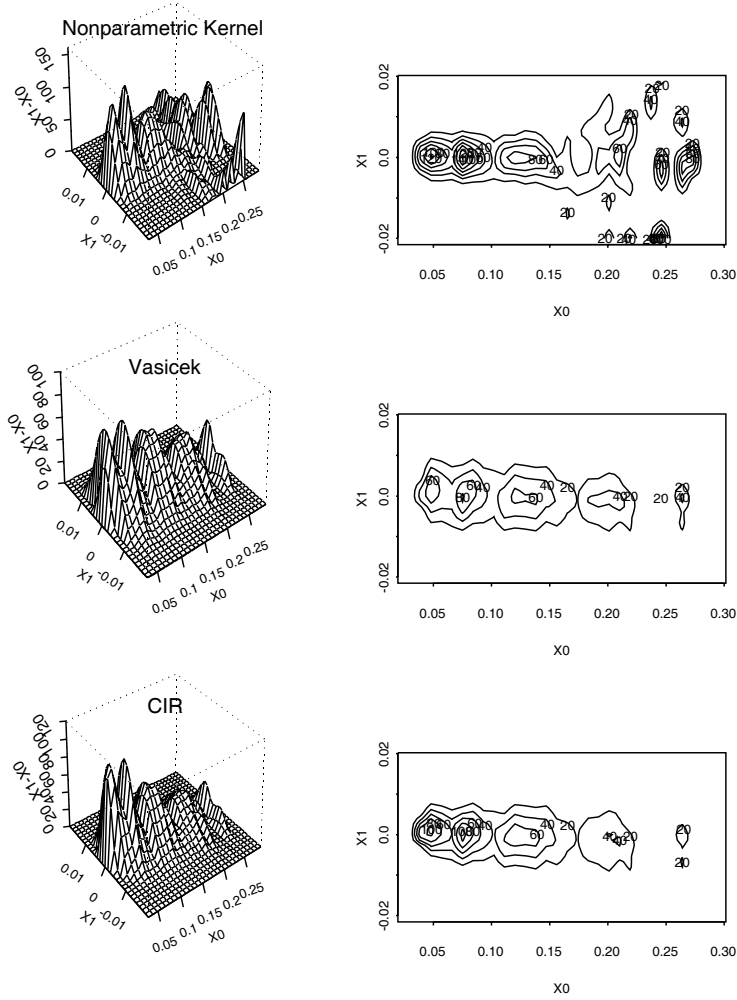


Figure 3 (a). Nonparametric kernel transtional density estimate and smoothed transtional densities for the Vasicek and CIR models at $h=0.007$. The right panels are the contours of the density surface exhibited in the left panels.

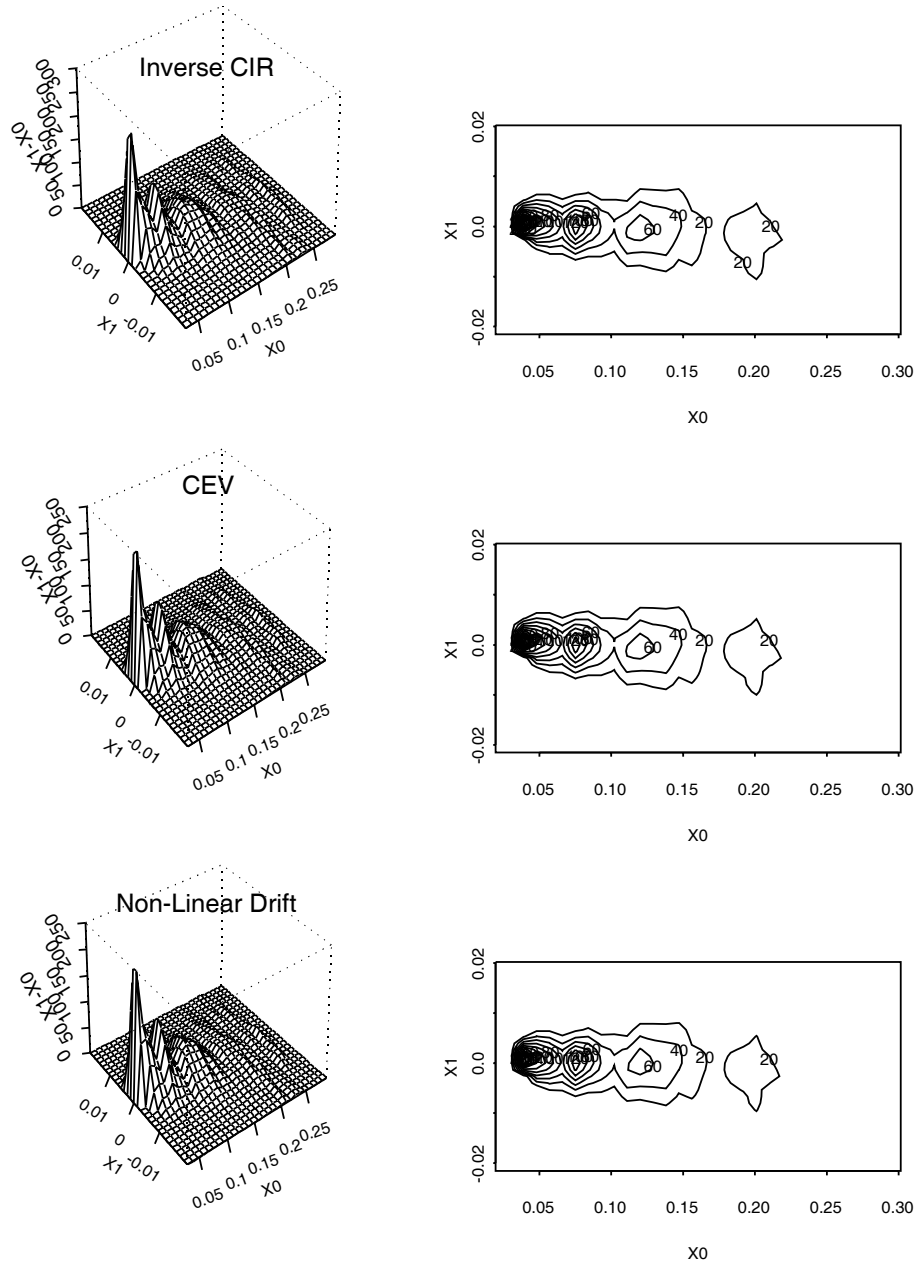


Figure 3 (b). Smoothed transitional densities of the inverse CIR, the CEV and the nonlinear drift models at $h=0.007$. The right panels are the contours of the density surface exhibited in the left panels.